Evaluation of the Evidence on the Effectiveness of Well Child Care Services for Children

Charles J. Homer, M.D., M.P.H.
Division of General and Ambulatory Pediatrics
Massachusetts General Hospital
Boston, Massachusetts

July 1988

Contractor Document
Health Program, Office of Technology Assessment
U.S. Congress, Washington, DC 20510

This paper was prepared by an outside contractor for the OTA assessment Healthy Children: Investing in the Future. The paper does not necessarily reflect the analytical findings of OTA, the assessment's advisory panel, or the Technology Assessment Board.

J 952-80

TABLE OF CONTENTS

	Page
I. Documentation of Recommended Well Child Interventions for Children From 1 month through 11 years of Age	1-5
II. Review and Critique of Studies Evaluating the Effectiveness of Well Child Care as A Whole	6
Introduction	6
Literature Review	11
Conclusions	28
III. The Effectiveness of 5 Well-Child Services	28
Physical Assessment	29
Overview	29
Literature Review	31
Conclusion	35
Denver Developmental Screening Test	36
Overview	36
Background, Reliability, and Concurrent Validity	37
Predictive Validity and Utility	39
Conclusion	43
Anticipatory Guidance for Injury Prevention in the Clinical Setting	44
Overview	44
Literature Review	45
Conclusions	47
Iron Deficiency (Anemia)	47
Auditory Screening in the Preschool Child	53
Overview and Introduction	53
Burden of Suffering and Effectiveness of Therapy	54
Screening Tests	56
Conclusions	60
IV. Cost Effectiveness Studies	60
V. Tables	66
VI. References	82

LIST OF TABLES

Table 1. Recommended Number of Well Child Visits 63

Table 2. General Summary of Physical and Developmental Evaluations Recommended for Child Health Supervision 64

Table 3. Recommended Performance of Specified Screening Tests 65

Table 4. EPSDT Screening Recommendations for Selected States 66

Table 5. Summary of Literature Evaluation Effectiveness of Well Child Care as a Whole 67-73

Table 6. Effectiveness of the Physical Examination in Well Child Care 74-75

Table 7. Predictive Validity of the Denver Developmental Screening Test 76

Table 8. Effectiveness of Anticipatory Guidance on Infant Car Restraint Use 77-78

Table 9. Iron Deficiency, Cognitive, Development and Behavior 79-81

I. Documentation of Recommended Well Child Interventions for Children from 1 month through 11 years of Age

The frequency and timing of recommended well child visits, and the frequency and timing of recommended screening tests or procedures in well child care, vary substantially among Western nations, among states within the United States, and even across time from any particular recommending organization. (American Academy of Pediatrics, 1974; Strain, 1983; Canadian Pediatric Society, 1983; Great Britain, 1976; General Medical Services Committee, 1984; David Hall, 1987; Aidan Macfarlane, 1987; Canadian Task Force, 1979.)

The specific recommendations of a number of supervisory bodies[1] for certain well child care components are presented in Tables 1 through 4. These recommendations apply to children aged one month to eleven years. In general these guidelines demonstrate the following characteristics:

1. A trend in recommendations from the American Academy of Pediatrics towards increasing the number of visits over the past fifteen years;
2. A greater number of visits recommended in the United States than in Great Britain, although fewer than commonly provided in other Western European nations;

---

[1]. The American Academy of Pediatrics, Canadian Pediatric Society, Canadian Task Force on the Periodic Health Exam, three British groups (see text), and various state Early and Periodic Screening, Diagnosis, and Treatment (EPSDT) programs.

3. Recommendations for a more focussed physical examination than in the past;
4. Increased concern with identification of behavioral and developmental problems, coupled recently, especially in Great Britain, with increased recognition of the difficulties in reliably and validly identifying such problems;
5. Lack of consensus, especially apparent among the states, concerning the appropriate populations for screening procedures, the optimal age, and the frequency of use;

Several caveats must be borne in mind in making comparisons between American and British schedules. Infants and children in the United Kingdom do not routinely receive immunizations from their physician; therefore, immunization visits are not included in these guidelines. In England, from the time of the creation of the National Health Service until the mid 1970's, children were served by a "tripartite" service. Preventive child care services were provided through a network of child health clinics, staffed primarily by health visitors--nurses trained in child oriented community medicine--and by clinical medical officers (specially trained physicans). Health visitors also undertook home visiting; the extent to which this is still done is not clearly documented. Ill children were seen by general practitioners in their offices, although child health clinics provided some ill care as well. Hospitalized children and children with chronic diseases were taken care of by pediatricians.

Beginning with the report of a special committee convened in the mid-1970's and chaired by Sir S.D.M. Court, general practitioners were

encouraged to perform an increasing amount of child health supervision activities, and health visitors were increasingly attached to general practitioner's practices rather than district health authorities. General practitioners were also encouraged to develop some expertise in child health supervision. The Handbook of Preventive Care by The Royal College of General Practitioners was written to address this need. (General Medical Services Committee, 1984)

At the present time, the actual delivery of child health services seems to vary dramatically across different regions of England and Wales. The current Working Party on Developmental Surveillance in Childhood was specifically established "to review and comment upon current practice in Child Health Surveillance in the United Kingdom and to make recommendations for future practice." The guidelines noted in this report are the separate but largely concurring opinions of the Chair--Dr. David Hall--and Vice-Chair--Dr. Aidan Macfarlane--on what this group is likely to recommend.

In comparing recommendations within the North American continent, one must also consider the dramatically differing perspectives of the recommending bodies.

Concerning the varying state Early Periodic Screening, Diagnosis, and Treatment (EPSDT) recommendations, understanding the nature of the EPSDT program is crucial. Rosenbaum recently reviewed the strengths and limitations of EPSDT (Rosenbaum, 1986) In brief, EPSDT is a federally mandated but state administered program to provide early detection and treatment for medicaid eligible children. Because the expenses of the program are in some measure borne by the states, recommendations for frequency of visits and mandated screening measures

reflect not simply medical judgement but political and economic judgments of what each state is willing to spend for child health services for the poor.

While the tendency of the state EPSDT administrators might be to restrict services, the tendency of the Academy of Pediatrics is likely to be expansive. The American Academy of Pediatrics is an organization of pediatricians; while first and foremost seeking to advance the health of children, this body also represents the professional needs of its members. The Academy has several times sought to compile documentation for the efficacy of well child care and its schedule of care; failing this, the Academy has relied on expert judgement to formulate its recommendations. The Academy forthrightly admits the relatively arbitrary nature of its schedule, although its primary caveat is that the recommendations constitute a minimum standard for intact and functional families. Even within the pediatric community, these recommendations have stirred controversy.(Hoekelman, 1983) The Canadian Pediatric Society was strongly influenced by the AAP recommendations;(Marchessault, 1987) it also considered the Task Force's recommendations when formulating its standards.

The Canadian Task Force on the Periodic Health Examination represents an entirely different perspective on the provision of preventive care. This body, convened in 1976, was charged with the task of making recommendations for periodic health examination by the Deputy Ministers of Health of Canada. Although a broad variety of disciplines were represented on the Task Force, its membership consisted primarily of academic physicians committed to the field of clinical epidemiology.

In its major report, the Task Force reviewed 78 potentially preventable conditions, graded the evidence regarding the effectiveness of interventions for these conditions, and made recommendations with varying degrees of strength about including these interventions in periodic health exams. The Task Force then formulated these recommendations into a series of age-specific "health protection packages." Many of the recommendations made by the Task Force, especially concerning anticipatory guidance and developmental evaluation, were classified as level "C", i.e., "There is poor evidence regarding the inclusion of the condition in a periodic health examination, and recommendations may be made on other grounds." Immunizations and dental examinations were the only recommended child health interventions beyond the neonatal period receiving an "A" level recommendation, i.e., good evidence to support inclusion.

The existence of these wide variations in recommended well child services, although understandable in light of the differing health care systems and of the varied organizational perspectives, nonetheless bespeaks a paucity of good information supporting any one program of care over another when considered in light of child health outcomes. The information which exists is reviewed in the sections which follow.

II. Review and Critique of Studies Evaluating the Effectiveness of Well Child Care as A Whole

A. Introduction

1. General

The literature on the effectiveness of well child care has been reviewed repeatedly over the past thirty years.(Lewis, 1971; Yankauer, 1973; Casey, 1979; Shadish, 1982; Roughmann, 1985; Charney, 1986; Korsch, 1985) Although the body of literature on which these reviews have been based has not dramatically changed, these reviews have drawn dramatically different conclusions, ranging from profound doubts to ringing endorsements. This variation in inference most likely reflects more the political context of each review than the considered body of knowledge.

This review will focus primarily on the methodology of those studies that seek to evaluate the effectiveness of well child care, or, better stated, of "child health supervision." We will focus on studies of child health supervision as practiced primarily in the United States, although a few studies with some relevance from other nations will be cited. We will consider only studies that specifically examine a health services intervention or delivery program; studies examining changes in morbidity and mortality over time or locale alone will not be examined.

In addition to considering the methodology of these studies, we will also consider the reasonableness of the hypotheses considered and the completeness of the outcome measures. Well done studies with non-

plausible hypotheses or studies with restricted outcomes under consideration, no matter how well executed, will be of limited value.

Effectiveness studies evaluate how well programs that function in the "real world" achieve their goals; efficacy studies consider how well programs achieve their goals when performed in an optimal situation and manner. An efficacy study asks whether an intervention is conceivably worthwhile; an effectiveness study asks whether such an intervention will indeed work. This review will consider both, but will try to indicate into which category the study falls.

2. Goals of Child Health Supervision

In order to measure the effectiveness of child health supervision, we must first identify what the goals are for this activity. Perhaps surprisingly, little unanimity exists in answering this issue. Indeed, the most recent Academy of Pediatrics publication providing Guidelines for Child Health Supervision does not even include any mention of the overall goals for this activity.(American Academy of Pediatrics, 1985) Despite this lack of agreement, we will for this review consider three goals for child health supervision against which to measure effectiveness:

1. The prevention of premature mortality--death in infancy, childhood, adolescence, and early or middle adulthood.
2. The prevention or reduction of morbidity, defined as functional limitation due to physical or mental illness.
3. The promotion of the ability of the child eventually

> to fulfill his or her societally determined adult role; in Western society this means at least the ability to live independently and be self-supporting, and more broadly to participate in the full range of adult social interaction. Short-term indices for this outcome might include measures of maternal-infant interaction or developmental testing. Intermediate range criteria could include school performance and behavior problems. Long-term measures would involve job history, income, marital status, or criminal records.

Other goals for child health supervision have been proposed. These include the provision of support and reassurance to a family and the provision of a "medical home" for each child in the event that illness develops. Further empiric work is needed to assess the utility of these aspects of child health supervision to children and families before they can be considered goals or standards for outcome evaluation.

Although child health supervision visits occur throughout childhood, the great majority take place during the preschool period. This period will be emphasized in this report.

3. Role of Social Factors

Although improvement in health status, as broadly defined, is the appropriate goal for child health supervision, positing such goals presents a substantial risk of failure in judging the effectiveness of such a program. Health status of the population in general, and of

children in particular, is far more strongly determined by social and economic factors than by the nature of medical care.(Black, 1980; Brenner, 1984) This relationship between socioeconomic conditions and health outcomes is true in nations with health care delivery systems that provide service without respect to family income as well as in nations such as our own where income has traditionally determined health care resource utilization. Moreover, this relationship holds not only for the developmental outcomes indicated in goals three, but in physical health outcomes as well. Roghmann recently emphasized that resources expended to improve the health of children would have the greatest impact if directed at housing, income, and education.(Roughmann, 1985)

4. Summary

The results of this review can be briefly summarized.

1. No evidence exists to support an overall effect of child health supervision as now performed on childhood mortality or morbidity. On the other hand, the sample sizes used in prior studies have all been inadequate to identify even a large change in the frequency of mortality. Moreover, the measures of morbidity in most, if not all studies, have been poorly suited to the pediatric population.
2. Some evidence supports the contention that participation in comprehensive child health care can reduce the frequency of hospitalization for acute medical illnesses. However, the inferences that can

be drawn from this observation are limited. From a cost perspective, the decreased frequency of acute hospitalization is balanced by an increase in surgical admissions for "corrective" procedures. From a morbidity perspective, hospitalization rates in children vary so dramatically across geographic regions that hospitalization most likely reflects practice style rather than morbidity *per se*.

3. Few studies have adequately considered developmental/social functioning outcomes. The evidence that exists suggests that child health supervision as performed exerts little influence in this arena. Some studies with substantial limitations in generalizability or internal validity imply that modifications in the practice of child health supervision may allow this practice to exert a positive effect on some proximate measures of social functioning outcomes.

Studies examining the effectiveness of child health supervision were drawn from several overlapping areas of health services research. These included: examinations of the impact of varying schedules of well child care; studies of the effect of the then newly developed comprehensive care programs in the 1960's and early 1970's; evaluations of the federally mandated child health supervision program--EPSDT (Early and Periodic Screening, Diagnosis and Treatment); considerations of different modes of health service delivery or insurance programs; impact studies of health program

cutbacks in the 1980's; and, finally, evaluations of child health service activities specifically directed at improving developmental outcomes. Although none of these categories directly address the question of the overall effectiveness of well child care, each provides some insight into this area.

A detailed summary of each study is contained in Table 5. This table outlines the locale and years considered by the study, the characteristics of the study population, the study design, the sample size, the nature of the intervention examined, the outcome measures used, the results obtained, and comments concerning the threats to validity and limitations of generalizability. The text will in most instances highlight the conclusions and limitations from each body of literature, rather than repeat the information available in the table.

B. Literature Review

1. Evaluation of Periodicity

Two studies exist which specifically considered the impact of reducing the frequency of recommended visits for low risk children.(Gilbert, 1984; Hoekelman, 1975;) The more recent study is both methodologically and conceptually more advanced than the prior report, in that it more accurately reflects the population base, had a larger sample size, used blinded outcome assessment, and included important indicators of social functioning as well as physical outcome measures. Nonetheless, both studies failed to identify any ill health effects associated with a decrease in the frequency of scheduled well child visits.

The major caveat in interpreting both studies is that in both

reports, the group randomized to the lower number of scheduled visits nonetheless obtained an almost comparable number of visits during the first years of life. In the Hoekelman study, infants scheduled for fewer visits were seen three additional times by the office nurses for immunizations; informal advice and consultation was often obtained at that time. In the Gilbert report, primiparous women randomized to the lower frequency group made an average of 1.25 unscheduled well child visits in the first two years of life; at the same time, families randomized to the 10 visit group averaged only 7.63 visits. Thus, the average number of well child visits in the first two years of life was 6.19 for the low-frequency group and 7.89 for the high frequency group. Although one can reasonably argue that little harm in low risk women is likely to occur from reducing the recommended frequency of visits for low risk infants, one cannot infer that dramatic changes in actual utilization will occur! Moreover, neither study allows generalization to infants at higher social or biologic risk.

2. Evaluation of Comprehensive Care

The 1960's witnessed a dramatic expansion in health services provided to the poor, as well as ferment in the conceptualization of the optimal delivery of personal health services. These processes found concrete expression in the creation of a variety of "comprehensive care" programs for low income children. The precise character of these programs varied, but in most instances consisted of personal health services provided by a pediatrician in concert with a social worker and nurse; additional aspects often included availability of after hours consultation and continuity in provider

over time. Some programs additionally included augmented outreach activities, such as home visiting and case management. The payment required for use of these services is not documented in most of the evaluation reports; however, most were probably free.

Evaluations that examine the effectiveness of comprehensive care programs are not exactly congruent with the evaluation of child health supervision. Such programs certainly provided diagnosis and treatment of acute and chronic illness as well. Several of the evaluation studies did indicate that utilization of well child services increased with participation in comprehensive care programs; therefore, evaluations of the effectiveness of such programs have some bearing on the overall issue of child health supervision. However attributing any positive effect specifically to the child health supervision component, as opposed to, for example, continuity of providers, is simply not possible.

Unlike the evaluations of periodicity, these studies deal almost exclusively with populations at substantially higher risk of mortal, morbid, and adverse developmental outcomes than the general population. Failure to show improvements in outcomes may be due to factors outside the purview of child health supervision (such as housing). At the same time, the more frequent adverse outcomes in such populations make detection of any improvement statistically more likely with small sample sizes.

Comprehensive care program evaluations utilized ecologic, cross-sectional, and clinical trial (randomized and non-randomized) methods. These will be reviewed sequentially.

The two ecologic studies, (that is, those studies considering a

geographically defined population rather than an individual as the unit of observation) used dramatically different outcomes. Gordis's well known study examined the frequency of rheumatic fever in census tracts where comprehensive care programs were introduced; he found significantly fewer cases of rheumatic fever among residents of these tracts, when compared to the frequency of this condition in comparable tracts in the city.(Gordis, 1973) He strengthened this observation by medical review of cases of rheumatic fever, and attributed the decline to those cases preceded by symptomatic respiratory infection, a category likely to be sensitive to medical care. This study--like all others of its type--is subject to the ecologic fallacy, which is that the individual members of the community may not have in fact experienced the exposure attributed to them (use of comprehensive health services) simply by their residence. Moreover, this study may have little implication for child health supervision, in that it deals with services for ill children--those with pharyngitis--rather than with well child services.

A similar census tract based study in Rochester, using hospitalization rates as the outcome measure, did not find an overall effect of residence in the identified tract on lowered hospitalization rates; indeed, the rates were lower in the control region throughout the study period.(Klein, 1973) Although the investigators did find lower hospitalization rates among users of the comprehensive care program services, this outcome could easily be explained the selection bias of who chose to use services. Indeed, the authors comment "It appears that the low risk population of the target area became the initial health center users."

Two studies creatively sought to use school attendance as a marker for functional status as an outcome measure for gauging the effectiveness of participation in comprehensive care services. A study in Pittsburgh compared school attendance for enrolles in a Children's and Youth (C&Y) Project with non-enrolles, controlling for socioeconomic status by residence in a housing project.(Kaplan, 1972) The investigators found a small but statistically significant improvement in school attendance--3.2 days, 95% Confidence Interval of .94-5.5. If those who were motivated and oriented towards healthy behaviors preferentially enrolled for care in the C&Y Project, this selection bias of who chose to use the health center could explain this difference in outcome. A before-after study in Charlestown (Boston), also examined the effect of participation (measured by number of visits) in the then newly developed Bunker Hill Health Center on school attendance.(Moore, 1973) The investigators found no statistically significant effect of the level of health center utilization on school attendance, although there was a trend towards increased absenteeism with increased use. The relationship between health center use and absenteeism, however, is most likely confounded by prior health status; this is, sicker children necessarily used the health center more, and were also more likely to miss school. Moreover, a trend towards increased absenteeism existed during the course of the study for all school children.

Two non-randomized prospective studies examined enhanced outreach together with comprehensive care. One, based in Appalachia, prospectively followed matched pairs of infants from two comparable but geographically distinct communities.(Briscoe, 1980) The

intervention consisted of 7 nurse home visits to provide counselling, support, health education, and advocacy. This study is included because the home visiting program was integrally linked to well child services. Other studies of home visiting are not reviewed here.(See, for example, Olds, 1986) No differences existed in any measure of health status, although the measures of health status were crude; no social function outcomes were measured. No significant differences existed between the groups in health services utilization either; although a trend towards fewer outpatient visits was noted, the authors did not include the home visits with the total number of outpatient contacts.

The other prospective study compared the impact of intensive follow-up and home visits in the experimental group with "usual care" in the control group. (Rogers, 1974) This study used a quasi-randomization scheme by including alternate babies born at the Fort Defiance Indian reservation Hospital in the experimental and control groups. This study likewise showed no significant difference in mortality or morbidity, the latter including both illness experience and physical health measures. Developmental outcomes were not reported. No significant differences existed in utilization as well. The sample size, however, was inadequate to demonstrate even a 50% impact on mortality; sample size was adequate for detecting 25 to 50% differences in some of the morbidity measures. The lack of effect on utilization could be due to the opposing effects of outreach of improving case finding on the one hand and improving management on the other.

Two properly performed randomized clinical trials of

comprehensive care were reviewed. The evidence here is likewise equivocal. Gordis undertook an randomized clinical trial of comprehensive versus usual care in high social risk mothers in Baltimore.(Gordis, 1971) No differences were found in infant mortality, hospitalization or other utilization rates, height for weight standards, or immunization completeness. However, the sample size was inadequate, the morbidity measures extremely constricted, and no measures of social competence were included. Finally, all the services provided were free in both groups. Alpert and colleagues, in the most widely reported study of the effectiveness of comprehensive care for children, evaluated the Comprehensive Care Medical Program at Children's Hospital in Boston.(Alpert, 1976) No significant differences in any measure of morbidity (not detailed in the report) were identified. The overall frequency of outpatient visits and of hospitalizations were similar between experimental and control groups; different patterns of use were noted. Experimental group children made more health supervison visits and control children made more ill visits. Experimental group children underwent more surgery early in the study period, had fewer hospitalizations in year two and experienced comparable rates in year three. Mothers of participants in the comprehensive care program did report improved satisfaction in some areas. Even with these very limited findings, some questions about the generalizability and internal validity of the study exist (see Table 5).

Thus, although two studies of comprehensive care found some improvement in health outcomes (Gondis, 1973: Kaplan, 1972), and

although no study found harm from such programs, none conclusively demonstrated that comprehensive care in general, or child health supervision in particular, results in improvements in major mortality or morbidity outcomes for children. None of these studies examined the realm of development or social function.

3. EPSDT

The Early and Periodic Screening, Diagnosis and Treatment Program is a federally funded, state administered program which mandates screening of medicaid eligible infants and children and referral for definitive treatment.(Rosenbaum, 1986) Although not formally child health supervision, EPSDT programs in most cases follow guidelines similar to the American Academy of Pediatrics 1981 Guidelines for Health Supervision. Thus, evaluations of the effectiveness of EPSDT in improving the health of poor children should in large measure reflect on the effectiveness of child health supervision at large.

Unfortunately, the outcome measures used in the evaluation of EPSDT are difficult to interpret.(Irwin, 1982; Keller, 1983) The primary outcome used in these evaluations is the number of "abnormalities" detected in a screening, or the number of "referrals" made, i.e, the number of abnormalities which are deemed to merit a referral for treatment. The effectiveness of the program is then judged by the change in the number of referrals/abnormalities with either time in the program or number of repeat screens. Both studies do report a decline in abnormalities with time in the program; in Irwin's report this decline became apparent only after adjusting for a general trend towards increased case finding. Because no specific

information is provided about the precise nature of these abnormalities, their importance or remediability, or the eventual improvement with therapy, the importance of reduction in abnormalities/referrals is difficult to interpret.

Additional caveats in interpreting the effectiveness of EPSDT are brought to light by Reis's review of unpublished Department of Maternal and Child Health evaluation projects.(Reis, 1984) These projects demonstrate great variability in the proportion of the eligible population which is indeed screened and in the proportion of those screened who are identified as having a problem. Moreover, a recent article examining prior use of EPSDT by developmentally disabled children found EPSDT use did not result in earlier diagnosis or treatment for these children. (Meisels, 1988)

Thus, although two studies demonstrated a decrease in the frequency of referrals/abnormalities with repeat screening, the interpretation of these findings is uncertain and the relevance to child health supervision limited at best.

4. Alternative Health Service Delivery Systems and Insurance Programs

One possible approach to the evaluation of the effectiveness of child health supervision is to examine health outcomes for children in different systems of care--such as solo or group practice, fee for service or pre-paid, HMO type programs--and consider any differences in outcome in light of different patterns of care. Thus, if health outcomes are improved for children who use systems that provide more child health supervision, we could infer--within limits--that child health supervision is effective.

Two major studies--one of which was reanalyzed in a separate report--bear directly on this approach. The first was undertaken by Kessner and colleagues in the early 1970's in Washington, D.C.(Kessner, 1974) Kessner developed and then used a "tracer" system of measures to assess the quality of care. These tracers were intended to reflect short term outcomes that both had intrinsic health significance and were amenable to medical intervention. The outcomes used were iron deficiency anemia, visual disorders, middle ear infection (acute and chronic) and hearing loss. Kessner found that after controlling for social class differences in who used the different types of providers, there were no differences in any measure of "health status."

Dutton reanalyzed the Kessner data, asserting that the prior analysis was inadequate.(Dutton, 1980) She argued that the means for controlling for social class differences used by Kessner were insufficient. Moreover, she noted that in comparing prepaid care to all other care, Kessner inflated the value of "all other care" by lumping poor performers (solo practice) with better performers (hospital outpatient departments). After more carefully adjusting for social class differences, Dutton found that users of both prepaid programs and OPD clinics tended to have better health status measures than users of solo-practitioners. Statistically significant differences were observed when all measures were considered together and the OPD and prepaid groups combined and contrasted with the solo practice group.

If true, do these findings have any relevance to the assessment of the effectiveness of child health supervision? The answer to this

is uncertain. Kessner did observe that participants in the prepaid plans were far more likely to have well child checks than users of other services. However, the validity of this observation may be limited, in that for this analysis he lumped emergency room users and outpatient clinic users together, did not review private practice records, and reviewed only a very small number of charts. Dutton attributes the differences in outcome more to the characteristics of the practitioners themselves (more training), their means of reimbursement (salary versus fee-for- service), their specialization (pediatricians rather than general practitioners), and their functioning in group settings (with formal and informal peer review) than to the performance of well child care. Even so, the differences in outcomes that she identified were small.

Major limitations also exist in the generalizability of the results of this study. The study was not nationally representative, but was done in one city. Almost all the participants in the study were black; the solo practitioners were also disproportionately black.

Although the use of "tracer" conditions represented a major step forward in outcome evaluation in health services research, the appropriateness of these tracers and their amenability to medical intervention--particularly relating to ear disease--remains an open question. Moreover, as in so many other studies, no effort was made to examine developmental outcomes.

The other major study in this area which bears on child health issues is the Rand report on "The Effects of Cost Sharing in the Health of Children." (Valdez, 1986) In this study, 1844 children in

956 families, representing a random sample--with certain limitations--of the population of 6 geographically disparate communities, were randomly assigned to one of several health insurance plans that offered varying percentages of cost-sharing. Despite the many different plans offered, the authors of the report group together all the cost sharing plans and contrast them with free care for most analyses. Study subjects were followed for three to five years. The study used a wide variety of health outcome measures. In particular, they considered measures of physiologic health--anemia, hay fever, middle ear fluid, hearing loss, and visual acuity; limitations in daily activities; mental health perceptions; and general health perceptions. Although growth and developmental outcomes were also assessed, they have not been reported. No statistically significant differences were observed overall for any outcome. The only potentially clinically significant difference noted between those in the free care and those in the cost sharing group was the prevalence of anemia among poor children who were anemic at the start of the study.

As with Kessner's study, the relevance of these findings to the evaluation of child health supervision depends on whether the different groups experienced varying levels of well child care. Only partial results of the utilization data for children have been presented.(Leibowitz, 1985) These data were drawn primarily from one site--Dayton; the poorest site, South Carolina, was not analyzed. In general, the study results suggested that although children in cost sharing plans reduced use of well child services somewhat less than they reduced acute and chronic illness episodes (76% as many well-care

episodes, 72% as many acute care episodes, and 63% as many chronic care episodes in cost sharing compared to free care children), they nonetheless did lower use of well child services substantially.

Haggerty and Starfield separately and extensively critiqued the Rand study.(Haggerty, 1985; Starfield, 1985) These authors criticized both the substantial attrition (40%) of the initial study group and the small sample size--especially concerning the issue of functional limitations and concerning sub-groups such as the poor. The choice of outcome measures, while an advance over prior studies, still appeared inappropriate in the pediatric context. The functional limitations measure demonstrated too little variability to be useful. The physiologic outcome measures--with the exception of anemia--may not be responsive to medical therapy. Haggerty specifically noted the lack of consideration of social functioning ("new morbidity") outcomes. Both authors, and indeed the authors of the original report, noted that the nature of the cost sharing plans protected all families--and poor families in particular--from excessive health care cost burdens. Finally, Starfield noted that, although differences between outcomes for free care versus cost sharing among the poor were not statistically significant, "those in cost sharing plans were in worse health at the end of the experiment than those in the free plan on six of the eight health measures...."

Considering both the original reports and the critiques together, one can reasonably conclude that cost sharing reduces utilization of both preventive and illness related services and that this reduction is unlikely to affect adversely the physical/physiologic health of low risk populations. The data are consistent with the hypothesis that

cost sharing adversely affects some measures of health status among the poor, although this is far from definitive. The specific impact of reducing health services use on developmental/social functioning remains unexamined.

5. Impact of Service Cutbacks

In contrast to the expansion in services in the 1960's and early 1970's, the 1980's have witnessed reductions in programs providing services to populations in need. One study specifically examined the health impact of the reduction in services offered by a public, well child clinic in a rural county in Maryland.(Alexander, 1986) Specifically, the clinic no longer offered physical health assessments and exams, and reduced the frequency of offering immunizations. The study was cross sectional, comparing health status and utilization in the county which had reduced its services to these measures in another demographically matched county. No significant differences existed between the study and comparison county in maternal ratings of child health status. Care for those children previously followed in the public health clinic was apparently provided by private practitioners--one in particular. Therefore, although a particular program was discontinued, child health supervision activities in all likelihood continued.

6. Development/Social Functioning and Child Health Supervision

The studies that examined the global effectiveness of child health supervision or child health care at large, reviewed above, do

not consider behavioral and developmental outcomes to any significant extent. We did identify one study, however, which specifically examined how different styles of child health supervision--as practiced in clinical settings--influence these outcomes; likewise we identified a variety of studies that examined how special types of child health supervision might affect such outcomes.

Chamberlin conducted the effectiveness study in this arena, comparing the influence of pediatricians using different degrees of "teaching effort" in their well child care on a variety of maternal and child behavioral and developmental outcomes.(Chamberlin, 1980) In support of the effectiveness of the teaching effort, he did find a strong correlation between teaching effort (i.e., pediatrician involvement in parent education) and maternal knowledge and a small but significant correlation between teaching effort and the mother's self reported use of positive interaction with her child. On the other hand, he found that increased teaching also correlated with increased reported behavior problems; no correlation existed between teaching effort and formally measured developmental test results.

Several weaknesses in this study, unfortunately, limit any inferences which may be drawn. The sample was drawn from one city, and primarily represents middle class children. The class limitation of the study became more pronounced as the study progressed, with selective attrition of lower socioeconomic status subjects. The non-random allocation of families to practices and practitioners raises the possibility that those differences which were observed were more due to who chose the practitioner than to the practitioner's impact. On the other hand, the average characteristics of all the physicians

in a practice were inferred for all patients using any physician in the practice; this non-systematic inaccuracy would reduce the likelihood of finding any differences. Also, including both the intervention and the positive interaction with child together in regression models to predict developmental testing results may have unfairly minimized the impact of the intervention if its effect is mediated through improving this interaction.

The other studies examining the effect of well child care on developmental and behavioral outcomes are more appropriately considered efficacy studies. The most methodologically sophisticated of these studies was performed by Casey.(Casey, 1980) In this study, Casey personally used counselling oriented towards improving the affective interaction of mothers and infants during well child visits; the control group received "usual care" from the same pediatrician/investigator. After six months, the intervention group ranked higher on scales of maternal-infant interaction; no difference in Bayley developmental tests were noted. Like Chamberlin's study, Casey's work implies that well child care as usually performed does not substantially affect outcome in the realm of development of socially important competences. The longevity of the detected effect in the experimental group, its uniqueness to this particular investigator, and its long term importance, remain unclear.

Two other studies also present randomized trials of behavioral intervention, albeit less standardized.(Gutelius, 1977; Cullen, 1976) Both sets of investigators offered somewhat idiosyncratic behavior counselling to a random sample of the eligible population; Gutelius and colleagues additionally provided extensive social support (18, 12,

and 8 visits in the first three years of life), a "cognitive stimulation program," group counselling, and dietary counselling (including prolonged iron supplementation and daily meat intake). Cullen found fewer fears in the intervention group, little difference in developmental test results, and significant worsening among the intervention boys in their school performance and behavior. Gutelius's results were more positive, although in most cases not blindly assessed. Early IQ tests showed differences, although these differences diminished over time. Fewer behavior problems were noted after age five, although sample attrition may have biased these findings. A variety of measures of self-confidence also showed improved results in the experimental group. Unfortunately, even aside from the methodologic limitations of the Gutelius study, the extensive nature of the intervention--far broader than the current conceptualization of child health supervision--makes it incorrect to generalize from these results to gauging the effectiveness of child health supervision.

Thus, as noted in the summary, little evidence exists to support the notion that child health supervision as now practiced substantially influences the development of social competence; on the other hand, this arena has not been well examined. Both Casey's and, to a lesser extent, Gutelius's studies suggest that some impact is possible through modification of well child care. The particular importance of the measures thus far examined, and their duration of impact, has not been evaluated.

7. Conclusions

The literature evaluating the effectiveness of child health supervision is perhaps more remarkable for its limitations than for its findings. No evidence supports the contention that child health supervision as performed significantly influences mortality or morbidity or that the development of the child's social competence is enhanced through his or her participation in this endeavor. On the other hand, methodologically, sample sizes have been uniformly too small, and follow-up too brief, to identify mortality changes; the conceptualization of childhood morbidity until recently--and perhaps still--has been inadequate to examine health care effects, and most investigators have not even looked at developmentally oriented outcomes. For these reasons, we remain constrained by the need to use expert opinion and good intentions rather than scientific data to guide the nature of child health supervision. Given that participation in child health supervision does seem to provide substantial satisfaction to both parents and providers, the value of such opinion should not be overlooked.

III. The Effectiveness of Five Well-Child Care Services

Observational studies of ambulatory care reveal that a typical well-child visit takes approximately 10-12 minutes.(Reisinger, 1980) According to guidelines for the child health supervision visit, the health care provider should include in this encounter an initial or an interval medical history, a physical assessment, developmental and behavioral assessment, and the provision of anticipatory

guidance.(American Academy of Pediatrics, 1985) Following the visit with the health care provider a number of preventive screening measures may be undertaken.

We will review here the evidence for the effectiveness of several aspects of the well child visit. These include: 1) the physical assessment, 2) the most widely used formal instrument for developmental assessment--the Denver Developmental Screening Test, 3) anticipatory guidance for injury prevention, and two specific health screening endeavors--4) the detection of iron deficiency anemia and 5) the detection of hearing deficits.

The criteria used here for assessing the effectiveness of these services mirror those used by the Canadian Task Force on the Periodic Health Examination.(Canadian Task Force, 1979) Specifically, a prevention-oriented service can be recommended if a) the condition under consideration is important, that is, presents a substantial individual or societal burden of illness; b) the condition can be effectively treated, and c) the diagnostic maneuver is good, that is, the maneuver is sensitive, specific, safe, simple, cheap and acceptable. Compliance with recommended interventions by both the patient and other providers is also important in assessing large scale screening measures.

A. Physical Assessment

1. Overview

The physical examination may be considered a non-automated multiphasic screening test. The physical examination is a series of diagnostic tests intended to detect a variety of medical conditions.

Given that the physical examination is indeed a diagnostic test, studies examining its effectiveness should be assessed by the same criteria used for judging reports on other diagnostic tests, specifically: a) is there an independent blind comparison to a gold standard, b) does the sample in the study represent a spectrum of the conditions under consideration, c) is the setting for the study well described, d) is the reproducability of the test determined, e) is normal and abnormal sensibly defined, f) is the individual contribution of a part of the test to the overall cluster of tests clearly determined, g) are the tactics for carrying out the tests well described, and last, h) is the utility of the test determined, that is, are patients better off for having undergone it.(Sackett, 1985)

The literature that evaluates the physical examination in well child care fails completely in meeting these criteria. The studies considered here do not specify in advance the conditions which are being sought by the exam. None presents data concerning the validity of either positive or negative findings obtained on examination. Although the study settings are usually well described, the settings in most cases are schools rather than physicians' offices. Most studies do not consider the reproducability of the physical examination undertaken. (Sackett has recently summarized the literature which examines the reliability of the clinical exam and notes that many of the procedures in the examination are of poor reliability).(Sackett, 1985) Normality and abnormality are not specifically defined in most of these studies; a few studies arbitrarily classify findings as to their severity. The specific contribution of the physical examination to overall detection of disease is not usually considered; when it has

been, however, the physical examination often identifies conditions already known to the parents and thus adds little to what can be identified by history alone. Finally, no study truly considers whether the patient is better off for having been examined.

2. Literature Review

The specifics of the studies evaluated in the physical examination of childhood are listed in Table 6. Only one study directs its attention to the examination of the first year of life, when most physical exams take place.(Anderson, 1970) In this study, Anderson asked Connecticut pediatricians--of whom fewer than half agreed to participate--to record their well-child exams and to indicate how many of these exams resulted in the detection of abnormalities. Anderson found that only 1.9% of exams resulted in "significant" abnormalities, and over one-half of these abnormalities were already known to the parent. Although Anderson concludes that this low yield did not justify frequent exams, the data presented are insufficient to draw this conclusion. No confirmation of either the positive or negative findings found in these examinations took place. No measure of the utility of the exam to the patient (or the family) took place.

Two studies consider the effectiveness of the preschool physical examination--examination by the physician of a child age 4-1/2 prior to school entrance. O'Connell reviewed the charts of those infants entering kindergarten in Rochester, Minnesota, who were born and followed at the Mayo Clinic, and evaluated at the Clinic before entering school.(O'Connell, 1976) The generalizability of the study therefore is limited and selection and migration bias may have

influenced the findings. Nonetheless these investigators found that only 3.1% of the preschool examinations as recorded detected a previously unknown abnormality. The authors did not determine whether more abnormalities existed and were not detected, whether those detected had false positive findings, and whether these detections influenced the health of the children.

The study by Welch raises the question of the validity of the limited yield of physicians' examinations.(Welch, 1982) In this study of all children entering kindergarten in a community, investigators compared the results for height, weight, hearing, vision, blood pressure, and dentistry screening as performed by technicians and nurses in the school with written reports filled out by the students' private physicians as part of their preschool examination. Although the investigators did not report checking of the validity of the negative results, they did at least establish that the testers underwent training and that positive test results were "verified." The school-based screening program identified 33% of those screened as having some abnormality; only 30% of those identified with an abnormality had been previously identified by the physicians. Although this suggests that the use of a screening test is a better means for problem identification among preschoolers than the physician's examination, this inference is not fully justified. First, physicians often defer doing tests if they know the tests will be performed in school. Second, Welch's comparison is between the screening results and the physician's written school record, which may or may not reflect the physician's judgment and findings. Third, reports of abnormalities detected by the physician but not noted on screening were not included

in this report. Most importantly, the utility of establishing these diagnoses was not considered.

The next series of studies considers the usefulness of the routine physical examination for the school aged child by a school physician. Yankauer and colleagues examined a representative sample of Rochester, N.Y., first graders--excluding hearing, vision, and dental screening--and then subsequently re-examined these students over the ensuing four years. In the initial screening year, the authors found that 21% of the children had an "abnormality," but most (78%) of these conditions were already known and under care.(Yankauer, 1985) The likelihood that a condition was indeed under care was greater for those children who were examined before school by a private physician (OR=2.2, p=.03). Follow up exams were of even less obvious use; only 1 in every 251 exams resulted in a condition being diagnosed that was not already under care.(Yankauer, 1956) Finally, they noted that of the 163 conditions initially identified (in the cohort remaining at four years), 99 of the abnormalities remained four years later.(Yankauer, 1957) Only in the categories at ENT (ear, nose, and throat) and emotional problems did more than half of the originally noted problems resolve by ninth grade. Yankauer also examined whether abnormalities present in fourth grade and also present one or more years prior to fourth grade were "under care". He found that for both ENT and emotional conditions, more than half (58%) were <u>not</u> under care. The authors argue that the low frequency of treatment reflects the relatively recent onset these conditions. An equally plausible hypothesis is that parents recognize the self-resolving nature of many of these problems.

Despite their widespread citation in the child care literature, these studies have serious limitations. Like their colleagues studying preschoolers, this team did not assess the validity or the reliability of their exams. Also, they failed to consider any potential adverse effect from problem identification ("labelling"). Perhaps the most crucial caveat in making inferences about these studies to apply to the physical exam in well child care is that these investigators assumed that the children under study had a physician outside of school and were "under care." In the absence of well child services, it is possible that many of the conditions Yankauer noted as "under care" would indeed not have been previously identified or treated.

Grant reviewed his and his colleagues' experience in screening school aged children in El Paso, Texas, and compared the "yield" of screening tests administered by "paramedical" personnel with that of the physician performed physical exams.(Grant, 1973) He specifically excluded diagnoses of emotional problems, rashes, and acute illnesses from his report. In his sample--which spreads across the full age spectrum of students--13.4% of those screened had an abnormality; 71% of these were detected by screening measures. Again, the examinations were not validated nor was their precision determined. The overall usefulness of the identification was also not established.

Kohler screened all children in a given community in Sweden at age 7 with a physical examination, following hearing and vision testing.(Kohler, 1977) His results were quite similar to those of Grant, with roughly 15% of those screened being identified with a problem, and almost half of these problems related to visual acuity. The methodologic limitations of Grant's study remained in this study as

well--no validity testing, unknown reliability, and no utility measure.

DeAngelis reported a similar study with somewhat different results.(DeAngelis, 1983) Comparing physical exams by nurse practitioners with screening tests performed by trained aides, she found that screening measures identified more problems and used time more efficiently than did physical exams. However, the physical exam was more likely to reveal an abnormality per child contact with health provider, and little overlap existed between the diagnoses established by physical exam and those established by screening measures. The same methodological limitations remain in this study (i.e., lack of determining validity of findings, test reliability and test utility). Moreover, Deangelis includes a variety of minor or self-limited conditions as "abnormalities"; these diagnoses were specifically excluded by other authors.

3. Conclusion

This report reviews those studies examining the effectiveness of the general physical examination, with or without screening measures for hearing or vision defects. Although these studies, except for one, concluded the exam has little merit, they nonetheless did not meet established criteria for studies evaluating diagnostic tests. The most glaring weaknesses of these studies are a) their failure to establish the validity or the reliability of their findings, and b) their failure to examine the usefulness of the identification of physical abnormalities.

This report did not consider studies that examine the effectiveness of specific physical diagnosis procedures, such as the Ortalani maneuver for identification of congenital dysplasia of the hip, forward bending for

detection of scoliosis, patch testing for discovery of strabismus, or abdominal palpation for detection of tumors. Given the more focused nature of these examinations, their effectiveness should be easier to clarify.

B. The Denver Developmental Screening Test

1. Overview

Unlike the literature examining either the effectiveness of well-child care or the effectiveness of the physical examination, the studies which examine developmental screening demonstrate careful attention to issues of methodology. Even so, this literature only recently began to assess the impact of developmental screening in terms of its long-term affect on developmental outcome. This area will be the focus of this section.

Pediatricians and educators agree that school and behavior problems and learning disabilities constitute a major source of morbidity for contemporary youth.(Nader, 1975) Diagnostic and therapeutic measures for these youth consume a major portion of total societal expenditures for children.(Nader, 1981)

Given that these problems are a major source of morbidity and functional limitation, as well as expense, are there efficacious means of primary and secondary prevention? A variety of comprehensive "early intervention" programs, ranging from elaborate, complex programs such as the Perry Preschool Project and the Frank Porter Graham Abecederian Project to more straightforward programs such as Head Start, have shown substantial positive effects in social competence outcomes such as school performance and job stability.(Chamberlin, 1987; Ramey, 1982) Other

programs, including home visiting, have also shown some positive effect.(Larson, 1980; Olds, 1986)

Thus, the conditions encompassed within the "new morbidity" present a burden of suffering and efficacious interventions apparently exist. Major additional prerequisites for recommending a preventive intervention program include the availability of an acceptable diagnostic or screening instrument and, given that the intervention here is not provided by the physician, the likely compliance of the client and the service provider with the recommendation to obtain services.

2. The DDST--Background, Reliability, and Concurrent Validity

The most widely used and recommended developmental screening tool for use by child health personnel is the Denver Developmental Screening Test or one of its adaptations--the DDST-S or the Parents' Developmental Questionnaire (PDQ).(Casey, 1986; Frankenburg, 1983) These latter two tests allow more rapid developmental screening, and thus increase the acceptability of the test to both patient and provider; nonetheless as they have been validated primarily against the full Denver Developmental Screening Test, consideration here will be restricted to the full test.

Perhaps the first question in considering the DDST is whether any formal developmental testing is truly necessary? Stated differently, are either parental observations or intuitive judgments by pediatricians sufficiently predictive of subsequent problems so that formal testing is not required?

The answer to the second part of the question is clear--physicians are probably poor judges of development. In studies over two decades ago, Bierman and Korsch noted that pediatricians diagnosed only one third of

"mentally retarded" two-year-olds and consistently overestimated the IQs of retarded children.(Bierman, 1964; Korsch, 1961)

The validity of parent observations remains an open question. Nonetheless, the delay until school entrance of the diagnosis of many learning and behavioral problems suggests that parental observation alone is also inadequate.

The DDST was developed in 1967, and initially standardized on a presumably normal, non-randomly selected group of children from Denver, Colorado.(Frankenburg, 1967) The scoring system for the Denver test was revised in 1971.(Frankenburg, 1971) The great majority of the validation studies for the DDST examine concurrent validity; that is, these studies compare the DDST to a more established test of developmental or psychological function. The most commonly used criterion tests for comparison are the Bayley Scales of Infant Development or the Stanford Binet Intelligence Tests.[1]

The initial concurrent validation studies of the Denver showed a sensitivity of 68% and a specificity of 92%, comparing abnormal results on the DDST with IQ or developmental quotient of less than 70 on the other criterion tests.(Frankenburg, 1971) A careful evaluation of the DDST more recently in a rural North Carolina county showed comparable results, with

---

[1]. The Bayley Scales of Infant Development were published by Nancy Bayley and the Psychological Corporation in 1969. These scales are the most widely used and extensively standardized measures of infant development. The instrument consists of two scales, a 163-item mental development scale and an 81-item motor scale. Kessen (in Levine MD, Carey WB, Crocker, AC, and Gross RT, Developmental and Behavioral Pediatrics Saunders, Philadelphia, 1983) notes that the predictive validity of the BSID are "quite modest" in the general population, but work better in "high risk" groups. Extreme values are also more meaningful than variations within the range of normal "... infants with unusually low BSID indices are more likely to become children with low IQs or motor dysfunction...."

a sensitivity of 68%, and a specificity of 95%, comparing the DDST to the Stanford Binet.(Sturner, 1985) Other studies have found lower sensitivities more in the range of 36% to 40%, but with comparably high specificity.(Harper, 1983; Appelbaum, 1978)

The reliability of the Denver--when administered by trained individuals under constant supervision--is quite good. (Pediatricians are usually not formally trained in the administration of the Denver and supervision is in most cases non-existent). Short-term test-retest reliability was 97% in the initial reliability study. Inter-observer agreement was also 97%.(Frankenburg, 1971) Although other reports have not been quite so glowing, the results are still acceptable. In a recent study, a kappa coefficient of .72 was obtained in comparing the results of different observers, indicating very good agreement.(Cadman, 1984)

Given these acceptable test characteristics, many authorities have recommended incorporating the DDST in some form into child health supervision.(American Academy of Pediatrics, 1974; Casey, 1986; Frankenburg, 1983) The primary purpose in doing so would be to identify children likely to have later problems, so that interventions could occur to prevent these problems. Additional reasons for administering the test include providing assurance to the parent concerning the normalcy of development and to identify and reinforce the strengths of the child in the parents' eyes. These additional reasons, while perhaps salutary, seem nonetheless secondary in importance.

3. The DDST--Predictive Validity and Utility

Three studies have been performed to evaluate the usefulness of the Denver in predicting global school performance (Table 7). The later two

are of excellent methodologic merit(Sturner, 1985; Cadman, 1987), the earliest is substantially flawed.(Camp, 1977)

The first study examining the ability of the Denver to predict subsequent school performance was performed by Camp and colleagues, using one of the original Denver validation populations.(Camp, 1977) This population consisted of some users of a neighborhood health center in Denver; these users were initially administered the DDST. If the test was abnormal, they were invited back for testing with the Stanford Binet; if abnormal, a certain proportion (not specified) was asked to return. Age criteria were then applied to this group who took both the DDST and the Stanford-Binet. Finally, school records were obtained for those members of this study group who were enrolled in the Denver public schools in 1972. School failure was determined by the child's placement in a special or below age level class, by achievement test scores more than 1.5 years behind grade level, by teacher report of significant behavior problem, or by IQ less than 80. The information concerning outcomes was gathered by individuals with no knowledge of screening test results; whether school teachers were ignorant of the screening test results was not indicated.

These authors found that the presence of an abnormal or questionable test result had a sensitivity of 78% and a specificity of 60% of detection of school failure. The prevalence of school failure in their final study population was 57%; this, however, was certainly not a representative sample of any population.

The drawbacks to the study are readily apparent. The representative nature of the initial population is not clear; certainly after the selective inclusions and exclusions, the study population is far from representative. The substantial drop out rate also presents a potential

bias. These biases may influence the validity of the estimates of the test's sensitivity and specificity; they necessarily influence the estimates of the test's predictive value, simply by increasing the prevalence of school problems. The high positive predictive value (80%) for the DDST vis-a-vis school performance in this study is therefore meaningless in most clinical settings.

Sturner and colleagues reported a more sound investigation of the predictive validity of the Denver, initially screening all children registering for kindergarten in Person County, North Carolina.(Sturner, 1985) Although these investigators again preferentially included those with abnormal Denver screening results in their follow-up group, the preferences were allotted with known probabilities, and therefore the investigators could weight their final results to come up with inferences about the county as a whole. School failure for this investigation was defined as either being in a special class or achieving less than the twentieth percentile for second grade on the California Achievement Test.

The North Carolina group found that 57% of those with school failure were previously scored abnormal or questionable on the Denver. while 87% of children not failing by second grade had a normal DDST result. When these investigators considered a two-stage Denver test, i.e., prescreening with a shortened Denver instrument, the sensitivity decreased to 26%, while the specificity improved to 94%.

Cadman and colleagues not only sought to clarify the predictive validity of the Denver in a community setting, but uniquely sought to determine the utility of a community based screening and referral program. (Cadman, 1987) All children registering for kindergarten, excluding those "children already known to have problems of development severe enough to

preclude their attendance in a regular classroom...," were randomized to receive either the DDST with counselling and referral, the DDST alone, or no developmental test. The excluded group was estimated to be roughly 3-5% of all children starting school. In Cadman's study, trained public health nurses administered the full Denver as the initial screening test. Children with abnormal, questionable, or untestable results (AQU) were re-tested two weeks later at home, again by trained nurses. Children with two AQU results were considered "positive." A broad array of follow-up measures were obtained, including special class placement, teacher ratings of learning or behavior problems, parent ratings of child well being, intelligence (WISC-R) and achievement (Wide Range Achievement Test of arithmetic, spelling and reading). Potentially significant attrition and refusal to participate in the study did occur.[1] Cadman's group found that a "positive" DDST, administered in two stages at the time of kindergarten registration, detected only 6% of those children who would not be in regular second grade class; over 99% of those who were in the regular 2nd grade class had received a normal screening result. Only 9% of students overall were classified as "school failures" by Cadman's criteria.

When those children who obtained a "positive" test result and received counselling and referral were compared with those children with

---

[1]. 11.5% of those with AQU Denver results on the second test were lost to follow-up, with equal amounts from the counselling and noncounselling groups. Of the random sample of those with normal tests invited to take detailed psychologic testing, 32% refused. Of the sample of those not undergoing Denver tests who were invited to undergo psychologic testing, 41% refused. Those who dropped out from the control groups appeared to have more teacher-judged learning problems than those who underwent testing. These characteristics might cause the study to exaggerate the true differences between those with normal and abnormal Denver results, but should not bias the estimate of the impact of counselling and referral.

similar results but who did not receive counselling and referral, no differences in any measure of school performance were noted. Parents of children who had received counselling reported significantly more parental worry about their children's school work than those without counselling. Children in the counselling group did receive more community based services. Even so, only 10 of the 28 children in the counselling group were actually seen by their community physician and referred for more definitive evaluation. Due to the low frequency of AQU results (1.3%), relatively few children--25 and 21--were included in the study and control groups respectively.

The major weakness of the Cadman study, aside from the small final sample size, is the brief period of time between the screening test and school entrance. This short lag time did not leave much time for pre-school intervention programs, and, once school began, the control group rapidly received school-based services. Earlier screening may have also allowed earlier identification of some of those excluded from the study group on the basis of "known problems of development," and potentially some of these problems may have benefited from intervention.

4. Conclusion

The Denver Developmental Screening test, when administered before school, has fair predictive ability. Assuming a 25% rate of school failure, the clinician can predict that a positive preschool screenee will fail in early school between 60 and 70% of the time; a negative screenee will avoid failure 76 to 89% of the time. If the prevalence of school failure is only 10%, however, a positive screenee will fail only one-third to one half of the time, while a negative screenee will be in a normal

program with a probability of 95%. Less encouraging to the prospect of reducing school failure, however, is the finding that from 50 to 95% of those who will experience school failure will be missed in a preschool screen. Moreover, the very limited evidence presented to date does not support the assumption that detection of a problem will result in improvement; indeed, the parents of these children seem to worry more with no improvement in outcome!

Administration of the Denver Developmental Screening Test at the immediate pre-school period therefore cannot be recommended at this point as a component of an effective means of reducing school failure. This report does not review whether earlier identification of developmental delay through the use of the DDST is a useful effort. No specific studies on this question were identified. Although the early intervention literature, noted above, is encouraging, eligibility for participation in such programs is in most cases determined by family socioeconomic and demographic characteristics, rather than by child developmental scores. If the use of the Denver, or comparable tests, is to be recommended in the context of child health supervision, this recommendation must come on intuitive or philosophic rather than scientific grounds.

III. C. Anticipatory Guidance for Injury Prevention in the Clinical Setting

1. Overview

As noted above, the well child visit in pediatric practice consists of an initial or interval history, the physical examination, behavioral and developmental assessment, anticipatory guidance, and, at

times, additional tests. Anticipatory guidance involves providing health education, information, or counselling in order to influence the parents' or child's behavior and thus favorably influence the child's health. Subjects considered in anticipatory guidance range from traditional medical guidance (such as avoidance of contact with children with certain communicable diseases) and nutritional advice, to suggestions for appropriate behavioral management at specific developmental ages and information concerning health behaviors such as smoking and alcohol use.

Medical practitioners traditionally spend relatively little time in such activities. Reisinger and Bires found that pediatricians spent 8.4% of the time of a well visit on anticipatory guidance activities, and that the proportion of time spent on such guidance diminished with increasing patient age.(Reisinger, 1980)

The provision of anticipatory guidance on injury prevention in general and on use of child restraints in particular offers an excellent opportunity to evaluate the effectiveness of this aspect of well child care. First, the outcome measure--proper use of child restraints--is objective. Second, the scientific underpinnings of the recommendation are clear--proper use of an approved child safety restraint will almost certainly reduce the child's likelihood of death or injury due to motor vehicle accident. The same degree of certainty does not exist regarding the value of advice about the precise timing and order of introduction of solid foods for infants or about means for preventing or modifying behavioral problems.

2. Literature Review

The studies evaluating the effectiveness of anticipatory guidance

in increasing infant and child restraint use are listed in table 7. Most of these studies have been reviewed by Bergman, Pless and Christopherson.(Christophersen, 1986; Bergman, 1982; Pless, 1978) As these reviews indicate, the early studies suffer from severe methodologic limitations.(Bass, 1964; Kanthor, 1976; Allen, 1976; Scherz, 1976) These limitations consist primarily of non-random assignment of experimental and control groups and use of parental self-report in the assessment of outcome (restraint use). Even if the results of these studies were valid, they may not be easily generalized to the population in that these studies rely in most cases on either military samples or white, middle class populations.

The more methodologically sophisticated studies failed to demonstrate a substantial effect due to educational interventions.(Miller, 1977; Reisenger, 1978; Reisinger, 1981) The most sophisticated study, and the one most relevant to the process of well child care as usually practiced, was performed by Reisinger and Williams.(Reisinger, 1981) These investigators offered an educational program, with written materials, verbal reinforcement, and direct demonstration performed by the physician, to parents using a pediatric practice during certain time periods. A comparison group, using the same practice at different periods, were not offered the intervention. Use of restraint was assessed by direct observation on entrance to the practice parking lot. Reisinger and Williams did find a significant increase in proper restraint use at two months; by four months, however, no treatment effect was identified, primarily due to increased use of restraints in the control group. Thus, the findings indicate that pediatricians can accelerate use of infant restraints in those

likely to use such restraints eventually.

3. Conclusions

Whether the limited efficacy of physician counselling in increasing proper use of infant restraints can be generalized to all of anticipatory guidance as usually performed is not certain. On the one hand, use of infant restraints requires initial expenditure of funds and requires repeated individual action. Moreover, as the recent repeal of seat belt laws in Massachusetts indicates, Americans maintain strong emotions about how to travel in an automobile. Furthermore, the health care provider may not be viewed as an authority in the area of child automobile safety. On the other hand, strong feelings often exist in families about most aspects of child rearing, such as disciplining and diet, and in these areas, physician expertise rests on a much more narrow body of evidence. Thus, the impact of infant restraint advice on restraint use may accurately reflect the generally limited impact of medically provided anticipatory guidance.

The limited efficacy of such guidance has stimulated the development of alternative approaches to traditional well child care. Osborne has proposed and piloted the use of group well child care. (Osborne, 1981) One evaluation of the impact of such care on safety behaviors--in particular concerning adjusting maximum hot water temperature in a home--showed a significant positive effect from this style of care.(Thomas, 1984)

III. 4. Iron Deficiency (Anemia)

The prevalence of anemia in a population has long been used as a measure of that population's health status, socioeconomic status, and quality of medical care.(Foxman, 1983; Kessner, 1973) The complexity

of the issues concerning screening for anemia, however, requires that the effectiveness of screening for this "condition" be reviewed. As before, criteria for judging the need for a screening intervention are the burden of illness presented by the condition, the effectiveness of intervention, and the quality of the test. Ideally, the assessment of the quality of the test should include an assessment of the test's overall utility. For the case of screening for iron deficiency, none of these criteria are now satisfied.

Anemia, itself, is not a disease. Rather, anemia is a laboratory defined entity characterized by a level of hemoglobin (Hgb--the chemical in the blood which carries oxygen) below an established cutoff. Although a wide variety of medical illnesses and nutritional deficiencies can cause anemia, in unselected populations of children, iron deficiency is the overwhelmingly predominant etiology.(Foxman, 1983) Indeed, screening for anemia in infancy and childhood is recommended, in large measure, as a screen for iron deficiency.

Does iron deficiency in childhood present a sufficient burden of illness to recommend a screening program? Using the currently accepted criterion for definition of anemia in infants 12-24 months old of 11 gm Hgb/dl (grams of hemoglobin per deciliter, or 10 milliliters) whole blood, from 10 to 40% of infants in this age group--depending on race and socioeconomic status--are anemic.(Hutton, 1985) Thus, the condition--at least in a mild form--is certainly prevalent. Severe anemia on the basis of iron deficiency, however, is now thought to be uncommon; a recent survey of children enrolled in a nutritional supplementation program in inner city New Haven found only 1.0 percent of children had a Hgb below 9.8 gm Hgb/dl.(Vazquez-Sloane, 1987)

Less well established, however, is the harm--the disease, disability, discomfort, and cost--caused by being iron deficient. The anemia, per se, is not thought to cause overt symptoms until the level of hemoglobin drops below 8 gm/dl.(Foxman, 1983) Moreover, two studies in children with hemoglobin disorders failed to identify a significant relationship between level of anemia and IQ.(Pollitt, 1976)

A large number of studies have sought to clarify the issue of whether "sideropenia" (i.e., iron deficiency itself) may adversely influence childhood behavior and performance.(Pollitt, 1976; Webb, 1983; Oski, 1978; Lozoff, 1982; Deinard, 1981; Lozoff, 1982; Walter, 1983; Deinard, 1986; Soemantri, 1985;) These studies are summarized in Table 9. Although a great deal of convincing evidence exists concerning the biochemical impact of iron deficiency, the results of these "clinical" studies are highly contradictory--in part due to different geographic settings, different age groups, different classifications of iron deficiency, different outcome measures, different duration of follow-up, varying attention to potential confounding factors, and inadequate analytic techniques. Although none of these studies show children with iron deficiency to be advantaged, an adverse effect of iron deficiency on learning and behavior remains "unproven."

Thus, concerning "burden of illness," mild iron deficiency is prevalent. Whether the condition is serious, however, remains uncertain. How effective is therapy for iron deficiency? No doubt exists as to the effectiveness of iron therapy in rapidly correcting the hematologic manifestations--i.e., the anemia-- and the biochemical markers of iron deficiency. For many children these manifestations

would also resolve without therapy, although resolution would occur more slowly. (Deinard, 1986)

As indicated in the table, three studies found that relative deficiencies in mental performance among iron deficient children did not improve with iron therapy (Lozoff, 1982; Deinard, 1986, Lozoff, 1987), while other studies did detect improvement with treatment (Pollitt, 1976; Oski, 1983; Walter, 1983). The negative studies attributed their failure to demonstrate improvement with therapy to inadequate duration of treatment, insufficient duration of exposure to iron deficiency to affect learning, or irreparable effect of iron deficiency at crucial periods on learning (Lozatt, 1987). An equally plausible reason is that iron deficiency is not the cause of the poorer performance. Whatever the true reason, these findings leave open the question of the efficacy of iron therapy for ameliorating the greatest potential burden of this condition.

Even assuming the seriousness of the condition, and the effectiveness of treatment, problems exist in establishing a screening criterion for iron deficiency. The first issue confronted in trying to establish a criteria for diagnosis of iron deficiency is defining the "gold standard" for diagnosis. The presently accepted standard for diagnosis of iron deficiency is a response of 1 gm/dl of hemoglobin to a therapeutic regimen of iron! (Soemantri, 1985; Reeves, 1984)

A large number of additional tests exist for identification of iron deficiency. Anemia itself is felt to be a late effect of iron deficiency. The earliest marker is the serum ferritin level, which is the storage form of iron. This test requires a venous blood sample (difficult for routine screening in children), and is expensive (charge

at Massachusetts General Hospital = $23.75). Another early marker is the serum iron level itself. This must be measured together with the transport protein for iron (transferrin) and is reported as the transferrin saturation level (proportion of transferrin used by the iron). Again, this requires venous blood, is expensive (MGH charge iron = $13.50, transferrin = $17.50), and is also variable depending on the time of day and the presence of either acute or chronic illness. A somewhat later marker--but still prior to anemia--is the erythrocyte protoporphyrin (EP). This is a measure of the precursor of hemoglobin, and is mildly increased in iron deficiency; this test is routinely done as a screen for lead poisoning, in which case it is markedly elevated. The EP can be done via fingerstick and is inexpensive and rapid. Although within laboratory reliability is good, inter-laboratory variability can be up to 15%.

The sensitivities and specificities of the different screening tests for iron deficiency have not been described in an unselected American population using the accepted "gold standard" for iron deficiency of response to a therapeutic trial. In a military sample of children with hemoglobin levels less than 11.5 Gm/dl, the lower (more commonly used) cutoff of 11 Gm demonstrated a sensitivity of 52% and a specificity of 66% when measured against this standard.(Driggers, 1981) Stated otherwise, the most commonly used criterion for iron deficiency anemia identified only half of those in this sample who in fact would respond to a therapeutic trial of iron, and incorrectly identified as iron deficient one-third of those who would not respond. Indeed, in this study, no single commonly used cutoff identified much more than 50% of those who responded to iron therapy.(Reeves, 1984) Kim and

colleagues described an elegant approach to choosing both a screening test and cut off values through the use of receiver operator curves; unfortunately, their study design resulted in over-representation of children with low (<10.5 Gm) hemoglobin levels, making generalization difficult.(Kim, 1984) In their sample, nonetheless, pre-treatment Hgb level per se was the best indicator (highest sensitivity and specificity) of subsequent response to iron; the EP also performed well, especially as a screening test for more severe iron deficiency (i.e., identifying those who respond to iron therapy with an increase in Hemoglobin of 2 Gm/dl).

Thus, although iron deficiency is both prevalent and easily treated, ongoing questions remain about the seriousness of the condition, about the response of the alleged complications of iron deficiency to iron therapy, and about the best means for detection. It is on this somewhat uncertain basis that recommendations exist for pursuing the diagnosis of iron deficiency. Reeves recently stated the dilemma:

> To date there have been no evidences of physiologic advantage from the common iron deficiency seen in childhood...On the other hand, since the deficit is small and may correct itself, saddling the healthy child with a protracted, perhaps difficult therapeutic course may not be justified.(Reeves, 1984)

Given the potential seriousness of the defects induced by iron deficiency and the ease of addressing the hematologic manifestations, continued early identification of high risk (low socioeconomic status, minority) infants

with either a capillary hemoglobin, hematocrit, or EP, appears reasonable, with a liberal threshold (e.g, Hgb <11.5, or EP> 35 micrograms/dl whole blood) for institution of a therapeutic trial of iron. Additional screening beyond infancy does not appear indicated based on current information. This entire field merits large scale studies of (a) screening tests and (b) impact of iron deficiency and therapy on intelligence, performance, and behavior, using representative populations and careful epidemiologic design and analysis.

III. E. Auditory Screening in the Preschool Child

1. Introduction and Overview

The American Academy of Pediatrics and other bodies concerned with hearing impaired children recommend a threefold approach to the early detection of children with hearing problems. (Committee on Standards of Child Health Care, 1977) This approach consists of identification of high risk newborns and infants through application of risk criteria[1] of infants and toddlers through monitoring of speech and language development (possibly including use of formal speech and language screening instruments), and of preschoolers through the use of some form of formal hearing screening test. This critique will focus on the effectiveness of hearing screening in preschoolers.

Understanding the rationale behind the threefold approach and its limitations requires some understanding of the process of hearing. In brief, sound waves are transmitted from the environment to the tympanic

---

[1] High risk infants are those with 1) Family history of childhood hearing impariment, 2) Congenital perinatal infection 3) Anatomic malformations of the head or neck 4) Birth weight <1500gm 5) Hyperbilirubinemia 6) Bacterial meningitis or 7) Severe asphyxia. (Joint Committee on Infant Hearing, 1982)

membrane (ear drum). The sound waves cause vibrations of the ear drum, which are transmitted via a series of fine bones through an air filled space, called the middle ear, to the inner ear. In the inner ear, a specialized structure called the cochlea converts these mechanical impulses to electrical impulses, which are then transmitted to the brain and interpreted as sounds.

Impairment at any step in this process will interfere with normal hearing. Thus, excess ear wax blocking the ear canal leading to the ear drum will diminish hearing. Fluid in the middle ear, either concurrent with or residual from an acute ear infection, may cause a "conductive" hearing loss. Damage to the inner ear structures, either as a result of congenital malformation, prenatal or postnatal infection, or the toxic effect of drugs or other chemicals (bilirubin) results in "sensorineural" hearing loss. Brain injury also may result in problems processing sound input. In general, sensorineural hearing deficits are more severe, bilateral, and permanent, while conductive loss is less severe, unilateral or bilateral, and transient.

B. Burden of Suffering and Effectiveness of Therapy

Severe sensorineural hearing loss probably occurs at a frequency of 1 to 2 per 1000 live births, resulting in 2000 to 4000 profoundly deaf babies being born each year.(Coplan, 1987; Black, 1975) Additional sensorineural loss results from postnatal infections such as meningitis or encephalitis, from certain medications often used in neonatal intensive care, and from elevated bilirubin levels which also occur most frequently in premature infants. This early and severe loss clearly results in impairment in both general learning and speech and language development.

Treatment for such severely impaired children includes use of hearing

amplification and a variety of special educational techniques. In addition, family counselling is often provided. Although the effectiveness of these interventions were not reviewed, the value of providing some therapy for such children and their families appears self-evident.

The frequency of conductive hearing loss is less clear cut. Again, the overwhelmingly dominant cause of conductive hearing loss in children is fluid in the middle ear (middle ear effusion), with or without acute infection.(Kohler, 1972) Middle ear dysfunction rates exhibit marked seasonality, with increased rates during winter months. Approximately 10-30% of preschool and early school age children will have detectable abnormalities of their middle ears; roughly half of these (5-10%) will have hearing impairment as a result of the presence of this middle ear fluid (i.e., half of those children with detectable middle ear fluid will not have detectable hearing impairment).(Kohler, 1972; Feldman, 1980; Fisch, 1981; Northern, 1984) Whether children with middle ear effusion experience long term problems in speech and language skills or are at increased risk for subsequent learning and behavioral disorders remains an open question. More severe, bilateral conductive losses, particularly at earlier stages of language development, most likely do cause short term speech and language delays.(Paradise, 1981; Ventry, 1980)

Most cases of middle ear effusion resolve spontaneously over a period of weeks to months.(Paradise, 1981) Antibiotics increase the rate of resolution, while decongestants do not.(Mandel, 1987; Cantekin, 1983) Surgical drainage, with or without the insertion of tubes, also resolves the effusion and any resultant hearing loss, although the duration hearing improvement may be brief.(Brown, 1978) Unfortunately, surgical

intervention also involves risk and expense, as well as uncertainty about the long term implications of tube placement.

The burden of disease caused by sensorineural impairment (1 or 2/1000 versus the estimated incidence of PKU of 1/10,000) and the likely effectiveness of therapy makes newborn screening an attractive possibility. Screening in the preschool period offers less attraction on a number of grounds: 1) age three to four is too late for optimal identification of congenital severe sensorineural deficits, in that critical language and learning periods are missed; 2) a great number of positive screens will likely be due to mild conductive hearing loss, with little attendant impairment or disability; and 3) the effectiveness of therapy is uncertain.

C. Screening Tests

Unfortunately, assessment of hearing on a screening basis is difficult in the newborn period and relatively easy in the preschool age group. As noted above, the AAP and the Joint Committee on Infant Hearing Screening--which includes representatives of the American Academy of Ophthalmology and Otolaryngology and the American Speech and Hearing Association as well as the AAP--recommend use of high risk screening criteria. Infants meeting one of these criteria are then referred for more extensive diagnostic audiologic evaluation. Alternative or complementary newborn screening tests measure the infant's behavioral (movement), physiologic (e.g., increased heart rate), or brain wave response to specific sound stimuli. The precise sensitivity and specificity of these tests, and their optimal combination for early identification of hearing deficits remain controversial. (Alberti, 1985; Pettigrew, 1986)

Preschoolers are in most cases screened through the use of pure-tone

audiometry. For this test, children are presented with sounds across a range of frequencies at a sound intensity (loudness) somewhat above the young adult normal level (i.e., 20 to 25dB) Older children may simply raise a finger or hand when they hear the sound; younger children may be conditioned to perform a more appealing task (e.g., put a peg in a hole) when they hear the sound.(Northern, 1975)

A variety of technical factors can interfere with the performance of pure tone audiometry. In particular, care must be exercised to maintain the machine in proper calibration (many newer machines are self-calibrating). Also, the testing environment must be relatively quiet, although complete soundproofing, necessary for complete diagnostic testing, is not required. Furthermore, other distractions to the child should be kept at a minimum. The extent to which practitioners observe these technical requirements has not been examined. Controversy also exists concerning the precise frequencies which should be tested, and the number of failures (one or two frequencies) which define test passage/failure.

The criterion test against which screening audiometry is judged is termed threshold audiometry. For this test, sounds across a range of frequencies are presented at either increasing or decreasing intensity to determine the "threshold" or minimum intensity which the individual perceives.

FitzZaland and Zink recently examined the sensitivity and specificity of pure tone audiometry in kindergarten and first grade students in a region of British Columbia.(FitzZland, 1984) These investigators compared results of pure tone audiometry (and other screening tests) with a "complete audiologic evaluation" which included threshold testing. The

authors did not indicate whether the audiologic assessment was performed blind to the screening results. These authors found a sensitivity of 93% and a specificity of 99%, with an overall prevalence of hearing impairment of 3.9%. Of the 137 children with hearing impairment, 10 had sensorineural defects and 4 had mixed conductive and sensorineural impairment; the remainder had pure conductive deficits.

Thus, pure-tone audiometry, properly performed, appears to be an excellent screening test for the detection of hearing impairment when judged against threshold audiometry. Whether this is the appropriate criterion has also been a subject of controversy. Some have argued that detecting deficits in speech perception, rather than sound perception, should be the goal of hearing screening. This perspective led to the development of the Verbal Auditory Screening Test for Children (VASC); unfortunately, validation of this test has been inadequate.(Northern, 1975)

An alternative perspective, popular in the audiologic and otologic communities, holds that pure tone audiometry is not sensitive enough for the detection of middle ear disease. Proponents of this perspective advocate use of impedance audiometry, which measures the mobility of the ear drum and assesses the level of pressure (positive, neutral, or negative) in the middle ear. Indeed, impedance audiometry is a sensitive, specific, and reliable measure of middle ear dysfunction, and pure tone audiometry detects only approximately 50% of cases of middle ear effusion.(Northern, 1984) However, the importance of the detection of middle ear fluid, in the absence of hearing impairment, remains entirely speculative.

As noted in the section on the Denver Developmental Screening Test,

optimal assessment of a screening test includes not only gauging the burden of disease, the characteristics of the test, and the effectiveness of treatment, but also judging the overall utility of screening--i.e., does offering screening improve the well being of the individuals and the community screened? Again, a Canadian group tried to assess the utility of preschool community screening.(Feldman, 1980) In this study, the authors compared the results of kindergarten hearing and vision screening in two communities, one of which had instituted pre-school screening. In the pre-school screened community, all children were tested 6 to 12 months prior to the kindergarten assessment. The pre-school screen had been performed with the VASC; the kindergarten screen was performed with pure tone audiometry. The study demonstrated little difference in the prevalence of abnormal audiometry test results at kindergarten (16.8% versus 14.1%) between the previously non-screened and screened communities. This finding was in marked contrast to the results for vision screening, which was associated with fifty per cent decrease in frequency in the pre-school screened community. Given the sample size (approximately 750 in each group) the study could have easily detected a 25% difference at the .05 level.

This study does exhibit several limitations. First, the authors do not indicate whether the kindergarten screener was blind to the results of the preschool screen. Also, although the authors tried to make the communities comparable through selection of relevant census tracts, unknown differences may have existed between the communities. Also, the investigators did not examine whether further ear evaluation or therapy was sought by those identified in the preschool screening, and whether the parents viewed these children differently as a result. Finally, the

actual outcomes for those individuals initially screened as abnormal were not examined. Nonetheless, the findings suggest that community pre school screening is not associated with a significant decrease in the prevalence of hearing deficits. The authors ascribe the failure of the screening program to the limited effectiveness of interventions for the treatment of conductive hearing disorders (middle ear effusion).

D. Conclusions

Issues surrounding the early identification of hearing deficits through screening in early childhood are surprisingly complex. Sensorineural deafness in infancy presents a significant burden of illness, particularly in terms of its developmental and social impact; unfortunately, infants are difficult to screen. Infants meeting the risk criteria outlined by the Joint Committee on Infant Hearing (roughly 7% of all infants (Lobavits, 1984)) should probably be referred for detailed audiologic testing. Conductive hearing deficits, while more prevalent, are in most cases less serious. Moreover, the efficacy of therapy is tentative. Pure tone audiometry, when properly performed, is a sensitive and specific means for detecting hearing deficits as measured by hearing threshold testing. Given the uncertain impact of most of these deficits, and the questionable efficacy of treatment whether pre-school children are better off for having been tested also remains unknown

IV. Cost-Effectiveness Studies:

Few of the studies reviewed in the sections above have demonstrated significant effectiveness of the interventions under consideration. In light of this, measures of the costs associated with the effectiveness are

understandably few! No concerted effort was made to identify studies specifically considering the costs associated with well child care.

Within the comprehensive care literature, Alpert considered cost.(Alpert, 1976) He and his colleagues noted that charges generated per illness were fewer for children in comprehensive care, largely due to lower rate of test ordering, compared to children using the emergency clinic.

Keller considered the cost-effectiveness of EPSDT in Michigan.(Keller, 1983) He did not find any decrease in unit cost with progressive number of screenings for an individual. He did note that costs for EPSDT participants were 7% lower than costs for non-participants, even accounting for administrative costs. Note, though, that non-participating EPSDT eligible children may be eligible for Medicaid due to large medical costs; thus, these may be non-comparable groups.

The Rand study, of course, contains substantial cost data.(Valdez, 1986; Starfield, 1985) Costs indeed were less for those who were assigned to the cost-sharing plans. Among younger children these lowered costs were largely due to decreased hospitalization rates; among older children, ambulatory services were more responsive to the cost sharing.

The costs associated with developmental screening were not characterized in the studies of the Denver Developmental Screening Test. One study did examine the cost of a paraprofessional administered screening program in a housing project.(Dawson, 1976) In this program, the cost was $16 per child screened, $32 per problem referred for diagnosis, and $67 per problem requiring treatment.

Studies specifically examining the cost-effectiveness of screening

for iron deficiency were not identified. Berwick and Komaroff presented a detailed cost effectiveness analysis of lead screening.(Berwick, 1982) They concluded that above a prevalence of 7%, and assuming both that mild elevations of body lead resulted in learning problems and that early treatment of elevated lead levels was effective, offering free erythrocyte porphyrin testing to three year old children averts morbidity and results in savings. Note that although the ability of the FEP to diagnose both lead elevation and iron deficiency is often touted as a reason for its use, the optimum time of screening for lead intoxication is three (when children are up and running), while the maximum risk period for iron deficiency is at approximately one year.

No further studies considering the costs associated with well child services were encountered.

Table 1. Recommended Number of Well Child Visits

	# of exams recommended			
Age:	1-6mos	7-12mos	1-4yrs	5-11yrs
American Academy of Pediatrics				
1974	4	1	4	3
1981	4	2	5	5
1985	4	2	5	4
Canadian Pediatrics Society				
1983	4	2	4	4
Canadian Task Force				
1979*	4	2	3	2
Court Committee, U.K.				
1976	1	1	3	not included
Royal College of General Practitioners,U.K.				
1984	1	1	2	not included
Draft Document, Working Group on Child Health Supervision, U.K., 1987**	1	1	2 ***	not included
Sweden, 1976@		4-7	4	---
Israel+		16 (5 MD, 11 nurse)	--	---
France++	6	6	9	---
Finland+++		12	3(Yr 2)	---

*visit numbers are for those recommended for the general population with A,B, or C level of recommendation

**Personal communications Drs. David Hall, Chair, and Aidan MacFarlane, Vice-Chair

***Full physical evaluation recommended at age 3 1/2; home assessment of walking and language use at 2 by nurse also recommended; screening tests for hearing and vision recommended before school entrance, with physical examination only if not performed at earlier time.

@3-4 home visits also occur in the first year of life; at age 4, a major "health control"--i.e., comprehensive physical, developmental/behavioral exam, is provided (Kohler, 1973)

+Palti, 1982

++Harris, 1974

+++Gilbert, 1984

BEST AVAILABLE COPY

Table 2. General Summary of Physical and Developmental Evaluations Recommended for Child Health Supervision, Ages 1 month through 11 years

	Physical Evaluation	Developmental Evaluation
American Academy of Pediatrics		
1981	"At each visit a complete physical examination is essential"	"By history and appropriate physical examination. If suspicious, by specific objective developmental testing"
1985	Specific evaluations recommended at each age. No mention of exam at 9, 12, and 15 month visits other than growth measurements.	Detailed developmental and behavioral guidelines provided at each age, with note of specific items for concern
Canadian Pediatric Society	Complete physical exam recommended at each visit. Specific items emphasized at particular times.	Behavioral history each exam. Language screening 7 times. School performance evaluation yearly beginning age 5
Canadian Task Force	specific physical exam measures recommended for most visits. Complete exams not recommended.	PDQ or DDST recommended most visits before age 2 1/2*; Review history of behavior problems ages 2 1/2,4,5,10. Assess parent-child interaction 18 months to 2 1/2.
Court Committee	Full examination at 6 weeks and pre-school; focussed exams at other times.	Review Development age 7 months, 18 months, 2 1/2, 4 1/2.
RC General Practitioners	Complete physical exam at first visit; brief exam thereafter. Specific points at each visit.	Milestone oriented developmental exam included in each visit
Working Group	Complete exam at 6 weeks, 8 months, and 3 1/2 years. Focussed measures at other times	Brief developmental assessment at 8 months. Home visit at 2 years with brief gross motor and verbal developmental evaluation. "grave doubts about the value of the neurodevelopmental exam"**

*Condition not reviewed by Task Force
**personal communication, Dr. David Hall

Table 3. Recommended Performance of Specified Screening Tests
Ages 1 month through 11 years

	Hearing Screening	Vision Screening	Hgb/Hct	Tb Testing	Urinalysis
AAP					
1981	4,5 yrs^	3-6,8 yrs.#	once ea. infancy, preschool, school	12 mos., then q. 1-2 yrs.	1 ea. period
1985	5 yrs^	3,6,8 yrs#	optional 9 mos.	high risk 9, 15 mos, 3-5 yrs.	5,7,9 yrs.
CPS	4,5 mos* 3,5 yrs*	6 mos., 3-6 yrs.	high risk 9 mos.	high risk 9 mos. 5 yrs.	not recommended
Task Force	2 1/2,5, 10 yrs.**	2-5 yrs.	low SES 9 mos.	high risk 5 yrs. (A), BCG age 5	not recommended
Court	7 mos. 4 1/2 yrs	7 mos.,2 1/2, 4 1/2 yrs.	not mentioned	not mentioned	not mentioned
RCGP	7 mos. 2 1/2 yrs.@ 4 1/2 yrs.	7 mos. 2 1/2 yrs 4 1/2 yrs	not mentioned	"BCG when appropriate"	not mentioned
Working Group	8 mos. 4 1/2 yrs.	4 1/2 yrs.	not mentioned	not mentioned	not mentioned

^subjective hearing assessment at all visits and hearing evaluation suggested with speech delay.
#subjective assessment at all visits.
*method not specified
**"clinical exam for hearing"--not clearly specified
@ STYCAR 5 toy test

Table 4. EPSDT Screening Recommendations for Selected States+

STATE	# VISITS	Tb Test	Hgb/Hct	FEP/Pb	U/A	Hearing	Vision
ALASKA	7	10m/5^	6m/6		5y/3	33m/3	*
CALIFORNIA	13	11m/2	7m/6	**	4y/3	3y/4	3y/4
COLORADO	15	9m/6	9m/3	12m/4	6m/9	--	--
CONNECTICUT	11	opt.	10m/2	opt.	opt.	6y/1	6y/1
D.C.	10	9m/5	12m/5	12m/1	2y/4#	--	--
GEORGIA	10	12m/7	6m/8	12m/5	1m/10	3y(HR)	*
ILLINOIS	11	10m/2	10m/5	15m/opt	15m/5	--	--
INDIANA	18	16m/2	9m/2	4m/10(HR)	20m/2	5y/1	"inspection"
IOWA	11	**	4m/7	(HR)	30m/5	*	42m/4
KANSAS$	10	15m/1	9m/4	--	30m/5	3y/4	3y/4
MASSACHUSETTS	17	1y/2	1y/3	1y/4	4y/2	5y/1	5y/4
MARYLAND	10	15m/1	9m/3	15m/2	2y/2#	9m/3	5y/2
MINNESOTA	9	13m/1	6m/4	**	2y/1#	*	*
NEW HAMPSHIRE	12	12m/2	9m/2	12m	12m/5#	*	*
NEW JERSEY	11	7m/3	7m/3	--	20m/2#	*	*
OREGON$	9	--	9m/6	--	--	--	4y/3
PENNSYLVANIA	17	**	**	**		19m/--	19m/--
TEXAS	9	13m/3	6m/2	6m/1	opt.	--	2.5y/4
WEST VIRGINIA	18	12m/2	12m/11	**	2m/17	4y/8	*

+states are included on basis of having sent periodicity schedule to OTA; sample is neither representative nor systematic
^age at first screening/total number of tests recommended through age 11.
*screening recommended at each visit. Test not specified; formal and informal screening not differentiated.
**"if indicated"
HR indicates for high risk groups only
# also recommends urine cultures
$ also recommends use of Denver Development Screening Test or Parent's Developmental Questionnaire.

TABLE 5. Summary of Literature Evaluating Effectiveness of Well Child Care As A Whole

Author	Years Data Collected	Study Population	Study Design	Sample Size	Intervention	Outcome Measures	Results	Comments
Frequency of Child Health Supervision Visits								
Gilbert	1979-80	Ontario Low risk	RCT	214 exp. 252 control	Decrease # well child visits from 10 to 5 in first 2 years	# Physical Abnormalities # Undetected Abnormalities Bayley HOME Maternal Anxiety Satisfaction with care	No Differences Detected	Small difference in actual number of well child visits--6.19 in exp. group and 7.89 in control
Hoekelman	1971-2	Rochester Low Biologic Risk--Clinic and Private	RCT	125 exp. 121 control	Decrease # well child visits from 6 to 3 in first year	Knowledge Satisfaction with care compliance utilization # undetected Abnormalities	No Differences	1. Extra visits occurred due to contact with nurses for immunizations. Extra visits scheduled for by staff for clinic patients randomized to low decreased visit schedule. 2. Inadequate measures of developmental/behavioral outcomes 3. Inadequate power to detect large differences in the frequency of physical abnormalities 4. Outcome assessment not blinded to study group

Author	Years Data Collected	Study Population	Study Design	Sample Size	Intervention	Outcome Measures	Results	Comments
Comprehensive Care Program								
Gordis	1967-70	Baltimore Primiparous <18 Years	RCT	120 exp 117 control	Comprehensive care (MD,RN,MSW--free) vs. usual care	Infant Mortality Hospitalization Clinic/ER visits Height/weight<10% % immuniz	No differences	1. Inadequate Power to detect large differences, except in ER use 2. Inadequate morbidity and developmental measures
Kaplan	1969-70	Pittsburgh Attendees 2 Schools in Low Income Neighborhood	Cross Sectional	525 exp. 700 control	Enrollment in C&Y Health Project--daytime program, peds,sw,rn, public health	School Attendance	Small, statistically significant difference with + effect of enrollment status (3.2 days)	Potential self-selection of healthier children into program (see text)
Moore	1968-71	Charlestown school-children undergoing complete PE	Cross sectional	991 total 3 groups	Degree of participation in health center--multidiscp., comprehensive, free PE	Change in Absenteeism	No significant change in absenteeism with participation. Trend to increased absence.	1. prior health status likely determined both utilization and absenteeism 2. Secular trend existed towards increased absenteeism
Alpert	1964-8	Boston, Children's Hospital--poor, no other MD, live near hospital	RCT	173 exp. 189 control	Comprehensive Medical Care Program- MD, RN, MSW; vs. usual care	Child Health Index Utilization Sickness and Drug Days Satisfaction Cost Process: Use of Preventive Services and Immunizations	No sig. difference any morbidity measure; similar frequency outpatient visit with more preventive visits; no sig. difference overall hospitalization--more surgical, fewer acute; improved satisfaction with wait and professional relationships; improved process measures.	1. Population=users of one hospital 2. Eligibility not clear 3. 30% dropout--probably not biasing 4. comparative nature exp. and control groups not documented 5. specific morbidity measures not noted in report 6. No developmental measures 7. Multiple comparisons for statistical testing 8. Introduction of medicaid may have minimized effect.

Author	Years Data Collected	Study Population	Study Design	Sample Size	Intervention	Outcome Measures	Results	Comments
Rogers	1970-2	Fort Defiance, AZ--Live born infants	Quasi-randomized Trial	116 exp. 119 control	Intensive follow-up and home visits vs. usual care	Infant Mortality Health Appraisal age 1: Uncorrected abnormalities, global health assessment, Hct, DDST (not reported); Hospitalizations, and outpatient visits	No significant differences	1. Inadequate power mortality analysis 2. Adequate power for some morbidity outcomes; appropriateness uncertain 3. Confounding of case finding and better care 4. No behavioral outcomes
Augustin	1970-1	NYC children enrolled in Montefiore-Morisania C&Y Project	Hybrid Design--first year enrollees compared to 2nd year	40 total	Not Described	number of illness visits to clinic during 2nd year of program participation compared to age matched first year enrollee Hospital Days per registrant	35% decrease outpatient visits; decrease in hospitalization rates from .36 to .102	1. No description population; 2. No description of program 3. Inadequate control group 4. Time of enrollment and acute needs related (confounded)
Gordis	1968-70	Baltimore Residents 5-14 yrs. in eligible census tracts	Ecologic (census tract as unit of analysis)	Not relevant 35,068 eligible Incidence 13.5/100000	Existence of Comprehensive Care Program in Tract	Rheumatic Fever Incidence (Rates)	60% decline ($p<.005$) in rheumatic fever rates in eligible census tracts	1. Ecologic fallacy (see text) 2. Not specifically related to child health supervision
Klein	1968-70	Rochester 1.catchment area residents; 2.health center users	1.Ecologic 2. Cross sectional	1.8000 exp. 7000 control 2. 1500 to 3300 users; 6000 to 4750 non-users	Comprehen-sive, multispe-cialty group practice 1. in Tract vs. not in Tract 2. users vs. non-users	Hospitalization rates and length of stay	1. Lower hospital admission rates and LOS in control tracts throughout study. 2. Users had lower hospitalization rates than non-users and lower LOS than non-users or control group.	1. Limitations in value of hospitalization rates as outcome. 2. Selection bias in use of health center
Briscoe	1975,1977	Hazard, Ky. Sample of children born at ARH hospital, matched to children born at comparable facility	Cohort study--exp. and control groups geugraphi-callly separate	65 pairs from 177 pairs in original group; 79 pairs in new study group	Home visits (7) for counselling, support, education, and advocacy, plus well child care	Health status: Physical exam, Otitis Media, Hct, Iron deficiency; utilization--admissions and outpatient/ER visits	No difference in health status measures; non-significant trend to decreased utilization in experimental group but home visits not included.	1. Inadequate control population (increased distance to MD for control group, better insurance for intervention group) 2. Inadequate power to detect differences in hospitalization 3. No behavioral outcomes

Author	Years Data Collected	Study Design	Study Population	Sample Size	Intervention	Outcomes	Results	Comments
EPSDT								
Irwin	1973-80	Before/After with separate controls for ea. time.	S.E. Penn. EPSDT eligible >18 mos. at 1st screen; screened at 2 yrs.	1831 children	Participation in EPSDT program	1. Identification of an abnormal condition requiring treatment 2. 8 treatable conditions identified standardized for number of conditions tested.	1. No difference in crude rates 2. When adjusted for secular trend of increased identification rates, rescreening was associated with a 26% decrease.	1. Rationale for adjustment for secular trends unclear 2. No specific information on importance of conditions 3. No individual health status measures
Keller	1979	1. Repeated prevalence 2. cross-section users vs. non-users	Michigan--pop. eligible for EPSDT entire year	1. 16,000 random sample 2. 10000 users; 6000 non-users	Participation in EPSDT Program	1. Referral Rates 2. Costs for participants vs. non-participants; with and without administrative costs.	1. Decreased referral rates with increased screening. 2. No consistent change in costs with increased numbers of screenings 3. Participants cost less than non-participants	1. Same criticisms as Irwin 2 and 3. 2. Non-participants are likely different than participants (selection bias)--e.g., non-screened Medicaid eligible may have "spent down" to get onto Medicaid roles.
Reis	NR		Review of unpublished MCH funded EPSDT evaluations	---	---	---	---	1. Great variability in proportion of eligible population screened (14-85%). 2. Variation in case finding rates (6-18%) 3. Although 50-80% of those identified with problems were treated, only 7-10% were judged to achieve maximum benefit 4. Large proportion of those diagnosed were not previously identified

Author	Years Data Collected	Study Design	Study Population	Sample Size	Intervention	Outcome Measures	Results	Comments
Alternative Health Delivery and Insurance Systems								
Valdez	1974-82	RCT	Random sample families from 6 communities. Some exclusions. 0-11yrs.	1811 children	Differing levels of health insurance	Physiologic Function: Anemia, middle ear fluid, Hearing loss, Visual Acuity Physical Health limitations in daily activity Mental and General Health Perception	1. Overall no significant difference in health measures with differing levels of insurance 2. Decreased utilization associated with cost sharing--Preventive services decreased by comparable amount to other services 3. For Poor children--if anemic at outset of study, 8% of those in free care anemic at end, while 22% of those in cost sharing were.	1. Sample attrition 30% 2. Plans not representative of those generally available to the poor. 3. Inadequate power for examination of role limitations and for subgroup analyses 4. Growth and developmental outcomes not reported
Kessner	1970-1	Cross Sectional	Washington, D.C. Random sample from specific neighborhoods. Predominantly Black. 6mos-11yrs.	1436 families--2780 children	Different types of providers: solo practice, ffs group, prepaid group, hospital OPD, Emergency Rm, public clinic	"tracer" conditions-- 1. middle ear infection/hearing loss 2. Fe deficiency anemia 3. Visual Disorders	Provider type had no significant influence on health status measures after controlling for SES Generally poor performance of preventive measures by providers--tests not done, abnormals not followed-up	1. Generalizability limited with 1 city, Black population, large N's inner city solo practitioners 2. ? of adequate controlling for SES 3. ? re: aggregation of provider types 4. Implications for preventive care uncertain; if valid, implication is that although prepaid programs provided more preventive care, outcomes no different.
Dutton	1970-1	re-analysis of Kessner data	Same as Kessner	same as Kessner	same as Kessner	same as Kessner	trend toward lower health status for users of solo practitioners relative to users of prepaid or OPD care. Lower satisfaction with OPD use.	1. ? re generalizability 2. Aggregate effect very small 3. ? appropriateness of linking OPD and prepaid care schemes

Author	Years Data Collected	Study Population	Study Design	Sample Size	Intervention	Outcomes	Results	Comments
Impact of Service Cutbacks								
Alexander	1983	Rural Md. Children born 1981 index and control counties	Cross sectional	322 index 242 control	Discontinued physical health assessment at public health clinic	Maternal Ratings of Health Status--global, growth, developeent, acute and chronic illness; use of well child services	No significant differences in health status Fewer incomplete immunizations in county where services stopped!	1. Many services remained available both through clinic and privately 2. Cross sectional analysis 3. based on maternal report 4. Sufficient power to detect 50% differences in most conditions; borderline for 25%.

Author	Years Data Collected	Study Population	Study Design	Sample Size	Intervention	Outcomes	Results	Comments
Developmental Outcome Studies								
Cullen	1964-73	Rural W. Australia other criteria not stated	Stratified, then randomized (RCT)	101 families 122 children ea. group	20-30 min. interview every 3 mos. in 1st year; then every 6 mos for 4 years. emphasis on gentleness, positive outlook,	1. Behavior symptoms 2. Family relations 3. Readiness for work 4. Basic Learning Ability 5. Early School Personality 6. Stan-Binet vocab 7. Describe a picture 8. Spontaneous speech 9. Draw A Man	Fewer fears, more school lateness; many behaviors with no differences; intervention boys worse in school related outcomes; no effect for girls.	Sample Uncertain Generalizability Uncertain Intervention not standardized Importance of outcomes unclear Plausibility of sex interaction limited
Gutelius	1963-76 (enrolled 65-9 with 6 year follow-up)	Urban Washington, D.C.; primigravid 15-18 year mothers with early prenatal care, IQ>70; no neonatal problems	RCT	47 exp. 48 control	Pediatrician and nurse well child visits in motor coach, 1 hour ea.; additional nurse visits--total 18/12/8 1st 3 yrs. Group counselling, medicinal Fe, cognitive stimulation progras.	Bayley Stanford-Binet WISC-R Behavior Profile School Readiness	Cognitive:Decreasing differences after age 3. Behavioral: improved social and self-confidence scores at age 3; fewer behavior problems age 5 on. Improved school completion by exp. mothers as progras evolved.	Generalizability limited due to idiosyncratic study population and intensity of progras; Outcome assessment not blinded. Intervention unstandardized. Late attrition in control group of better performers.
Chamberlin	1976-9	Rochester. primiparous mothers recruited from pediatri-cians.	Cohort	371 total	Classifica-tion of peds practices by pediatrician involvement in parent education	maternal: knowledge, attitudes, child-rearing style Child: Behavior, development	Increased knowledge with increased teaching; No effect on development; increased reported behavior problems; small but sig. correlation teaching and positive interaction	Middle class population; All providers in one practice given average rating (measurement error). Attrition to lower SES families. Regression technique may have masked study effect by including intervening variable. ?Selection bias.
Casey	1977-8	North Carolina; primiparous mothers, no medical complica-tions; no other MD child care	RCT (randomized after stratifi-cation)	15 exp. 17 control (of 59 eligible)	counselling emphasizing affective interaction; control of WCC by same MD.	8 scales Maternal-infant interaction; Bayley; Object Permanence and vocal imitation scales	all scales favored intervention; sig. differences 4/8. No sig diff. Bayley. Vocal imitation favored intervention p<.1	Short follow-up; outcome measures of uncertain significance. Power limited. Generalizability limited by population and perhaps nature of intervention (unique to provider?).

Table 6. Effectiveness of the Physical Examination in Well Child Care

Author	Years Data Collected	Sample	Method Data Collection	Validation	Reliability Assessment	Utility Assessment	Yield	Comments
Infant								
Andersen	1969	44% Practicing CT pediatricians 100 consecutive well child exams	Physician Report of Abnormality	None	None	None	11.4% exams resulted in abnormality; 1.9% in significant abnormality; 80% discovered by 6 mos.	Parents unaware of abnormalities needing rx 62% of time. Study of limited value
Pre-School								
O'Connell	1970	382 born Mayo clinic, underwent preschool exam and entered KS 1970	Chart Review	None	None	None	3.1% exams resulted in previously undetected abnormalities (serous OM, myopia, amblyopia, color blind, speech disorder)	Biases in sample selection
Welch	1978	1158 entering KG, Roanoke, VA 1977	Comparison of School screening program with written physician preschool report	Study in one sense is validation of prior physician exam; positive findings of screening were "confirmed"	Not clearly specified; screeners underwent training.	None	33% of children had abnormalities; 91% of these detected by screening; 30% detected by physician exam.	abnormalities detected by exam and not screened for are not discussed.

Author School Aged	Years Data Collected	Sample	Method Data Collection	Validation	Reliability Assessment	Utility Assessment	Yield	Comments
Vankauer (A)	1952-3	1036 1st grade children from representative sample of schools	Examined by 1 MD; vision, hearing and dental problems not included	Limited--if in doubt, a secon opinion was sought	None	See (C)	21% children had abnormality; 78% under care and 12% were known; if preschool family MD exam, condition more likely under care.	relies on adequacy of care by an outside (family) physician
Vankauer (B)	1952-6	617 of above remaining for 3 years and 284 remaining one or two years	same as (A)	same as (A)	None	See (C)	14% develop new condition, primarily emotional and ENT; 50% under care before school exam.	1/251 exams resulted in a condition diagnosed not already under treatment
Vankauer (C)	1952-6	same as (B)	same as (A)	same as (A)	None	Of 163 conditions initially ident'd, 99 still present in grade 4. Most new conditions also present grade 4. ENT and emotional probs most likely to improve; if active, these probs least likely to be in care.	see Utility assessment	no examination of "labelling"
Grant	1967-70	6058 students in El Paso schools undergoing annual screening, age 5-18 yrs.	paramedic screening tests; physician physical exam; rashes, acute illnesses, emotional problems excluded	None	None	None (authors judged a detected condition worthwhile even if referral resulted in a "diagnosis" of no significance, such as functional murmur)	13.4% had abnormality detected--9.5% by screening, 3.9% by exam.	5.3% untreated defects were problems of visual acuity
Kohler	1969-72	649 children age 7 in one town in Sweden	Author examined all students.	None	None	None	15% had abnormality detected; half were vision problems; half previously known. PE detected functionally important abnormality in 4.2%.	
DeAngelis	1980-1	12,997 rural students; little access to medical care; nurse practitioners;	aides administered screening tests. Nurse practitioner did physical exam	None	None	None	907 of 2691 students undergoing PE had a problem identified; only 17% previously known.	little overlap in conditions; acute, self limited problems included. No utility measure.

Table 7. Predictive Validity of the Denver Developmental Screening Test

Author	Years Data Collected	Sample Characteristics	Outcome Measures	Prevalence of School Failure	Sensitivity Abn+Quest	Specificity ABn+Quest
Camp	1969-72	Low Income Denver residents using a Neighborhood Health Center; took DDST; if abnormal, asked back; if normal, some asked back. Of these, those over 8 years old before 9/73 and still living in Denver int eh Public Schools were included. 493 initially came back; 92 met age criteria; follow up on 65 of 92.	Special Class or Repeat Achievement test >1.3 years behind Significant teacher rated behavior problem Diagnosis of hyperactivity IQ < 80	57% with either IQ below 80 or learning problem	78%	68%
Cadman	1980-4	All children registering for normal kindergarten in 3 or four regions of Niagara, Ontario. Children randomized to receive DDST with counselling; DDST without counselling, and no DDST. All abnormals and random sample of others underwent further testing.	Teacher and parent reported learning problems Child not in regular class parental worry WRAT WISC-R Child Well Being Questionnaire	9% not in regular 2nd grade class	6%	99%
Sturner	1978-80	All children registering for kindergarten in Person County, NC, screened with DDST-S; follow-up testing on differing proportions of abnormals (100%), questionables (50%) and normals (10%).	Special class or repeat CAT-R < 20%ile	27% not in regular class or <20%ile on CAT-R	57%-1 stage 84%-2 stage	87%-1 stage 94%-2 stage

Table 8. Effectiveness of Anticipatory Guidance on Child Restraint Use

Author	Year	Site/Practice Style	Sample Size	Allocation method	Intervention	Outcome Assessment	Results	Comments
Bass	1982-3	Pittsburgh/ private practice	1423	1) control group=users one practice 2) different experiential groups=users another practice at different times	1)Letter by MD 2)Letter by MD + counselling 3)Letter by safety organization	maternal report of seat belt installation, by phone.	19.6% no info 19.1% org. letter 15.3% MD letter 43% MD letter + counselling	concerns re: biases in allocation and assessment
Kanthar	1974-5	Rochester/ prepaid health plan	16 exptl. 19 control	Quasi-random (every other infant born)	1)counselling by MD + pamphlet at prenatal visit. Control = no education.	maternal report, occasionally verified	42% use no info 69% MD info (p=.21)	small sample size bias in assessment no sig. difference
Allen	1974-5	Seattle/ prepaid health plan	202 of 500 eligible	volunteers for non-concurrent intervention groups	1)informational material only 2)informational material + film presentation 3) informational material, film presentation, and rehearsal of car seat use. control=no info (postpartum)	maternal report--questionnaire	1)37% no info 2)54% info only 3)71% info +film only 4)60% info+film+ rehearsal	selection bias assessment bias not necessarily relevent to office practice
Scherz	1970-4	Tacoma/ military well child care	300	random allocation	1)no information 2)display 3)display+pamphlet 4) (3) + nurse counselling 5) (3) + MD counselling	maternal report--questionnaire at 8 weeks and 9-12 mos.	at 8wks/12 mos, %safe= 1) 9/77 2) 12/74 3) 8/75 4) 22/81 5) 13/80	bias in assessment due to military population.

Author	Year	Site/Practice Style	Sample Size	Allocation	Intervention	Outcome Assessment	Results	Comments
Miller and Plese	1975-6(?)	Rochester/ Pediatric group practice	654 (age 0-17)	Randomized	1)pamphlet+verbal info., 2)pamphlet+verbal-+slide/tape control=no education	maternal questionnaire, rough validation with direct observation	no significant differences between control either intervention group	power! not a "physician" intervention per se.
Reisinger and Williams	1976-7	Pittsburgh/ in hospital program	1107	consecutive time intervals (non-concurrent controls)	control=no educ. 1)literature only 2)literature+health educator 3)literature+free car seat	direct observation at hospital discharge and 2 mos. follow-up	very low use at time of hospital discharge, no study effect; gradient from control to free seat with use at 2 mos., i.e., 26%/31%/36%/41%. Only free group statistically significant different from control.	rates may be inflated compared to general population in that more educated parents both more likely to use seat belts and to come for follow-up..
Reisinger and Williams	1978-9	Pittsburgh/ private practice	869	non-concurrent intervention adn control periods	control=no info. study=education by pediatrician with discussion, pamphlet, and demonstration	direct observation at 1,2,4,9, 13 mos.	Significant difference at 2 mos. (50 vs. 29%); no difference from 4 mos. thereafter.	attrition ranged from 10-23%.
Kelly,Sein, McCarthy	1981-2	New Haven/hospital primary care center	109 (6 mos. old)	randomized	3 part, developmentally oriented course, safe home pictures	parental knowledge, home safety assessment (direct), reported auto practices and accidents	a)improved knowledge, b)fewer home hazards, c)no change in restraint use (decreased sitting in front), d)no change in auto accidents	primarily parental report of outcomes.

Table 9. Iron Deficiency, Cognitive Development, and Behavior

Author	Sample Population	Age	Study Design	Classification of Iron Deficiency	Outcome Measures	Results	Comments
Webb/Oski	inner city Philadelphia jr. high school; Black; 92 study/101 control	12-14 yrs.	Cross-sectional	Hgb < 11.5 vs. Hgb > 14	Iowa Basic Skills Achievement tests	poorer performance if anemic; increased risk among older anemic males.	no control for SES no measure of Fe status sex-age interaction implausible
Oski/Honig	Syracuse outpatient clinic; 24 iron deficient infants	9-26 months	RCT of intramuscular iron therapy	Hgb < 10.5 and MCV (red cell size) < 73 cu. micron and Fe < 50 and transferrin sat < 12%	Bayley Scales Infant Development (BSID) at baseline and 8 d after rx.	significant increase in Mental Development Index (MDI) in deficient group, inversely proportional to initial Hgb; deficient children become more "reactive" with iron	initial MDI values not included in ANOVA model examining study effect
Lozoff	Guatemala; community based; 24 experimental and 40 control	6-24 months	RCT of oral iron	Hgb > 12.0 control vs. Hgb <10.5 and 2 of following 3: transferrin sat < 10%; EP > 100/dl RBC's (^ to 35/dl whole blood)); and ferritin < 12.	BSID	a)pretreatment MDI lower in anemic children; only significant in 19-24 mos. group; MDI for non-anemic group=120. b)correlation (Pearson) between pretreatment MDI and increasing levels of iron sufficiency=.73. c)no significant improvement in MDI scores with 7 d oral iron therapy	a)power adequate (=.75) to detect 7.5 point difference in MDI scores b)anemic children also had other evidence of malnutrition c)no control in analyses for SES differences between anemic and other groups d)MDI for non-anemic group higher than would be expected.
Oski/Honig	Syracuse outpatient clinic: non-anemic with differing levels of iron deficiency; total n=38	9-12 mos.	Before/After intramuscular iron; (RCT not approved by IRB)	All Hgb >11.0; Depleted: Ferritin <12; Deficient (a): Depleted+EP >30; Deficient (b): Deficient (a)+MCV <70	BSID	a) iron depleted group comparable (low ferritin only) comparable to sufficient group on all outcomes. b) when aggregated by sufficient+depleted vs. deficient a+b groups, deficient group had lower pretreatment MDI (93.7 vs. 84.4, p=.175) and greater response to rx (5.9 vs. 21.6 pts, p=.01), ending with higher MDI scores (90.6 vs. 106, p=.2)	a)no SES control except matching by race b)inadequate analysis (fails to account for regression to mean by including initial values in ANOVA) c)implausible result of higher final MDI scores.

Deinard	Minneapolis and Duluth public clinic users; non-anemic with differing levels of iron deficiency; n=34(severe), 21(mod), 157 (control)	11-13 mos.	Cross-sectional	All Hct > 34; severe: ferritin < 9; moderate: ferritin 10-19 normal: ferritin >19.	BSID visual "habituation" Uzgaris and Hunt Ordinal Scales of Psychologic Development I,II,V (measures visual pursuit, means of obtaining ends, and construction of object relations in space)	No differences between groups on any measures	no separate SES control, although relatively homogeneous group no physiologic measures of iron deficiency (e.g., EP)
Walter	Santiago,Chile; infants in infant feeding study; normal birthweight and growth. Iron sufficient vs. iron deficient without anemia vs. iron deficient with anemia; total n=37.	15 mos.	Before/After oral iron for 15 d.	Anemia: Hgb<11 Iron deficiency/Anemic: Hgb <11 and 2 abnormal tests (transferrin <10%, EP>100, ferritin <10);Iron deficiency/ non-anemic: Hgb>11 and one of above criteria	BSID (including behavioral questionnaire)	a) Pretreatment mean MDI lower in anemic vs. other groups; no difference between non-anemic deficient and sufficient (98 vs. 108, p=.0025; vs. 113, NS) b) Significant improvement in MDI of anemic children with iron therapy (98 to 108, p=.01), no change in other groups except post-hoc observation that non-anemic with two findings of iron deficiency (n=6) increased from 108 to 118. c)Improvement in measures of cooperativeness and attentiveness in iron deficient/anemic group.	no SES measures inadequate analysis (regression to mean, as above)
Deinard	Minneapolis public clinics; normal birthweight and growth; anemia/iron deficient vs. non-anemic iron deficient vs. iron sufficient; total n=70	18-60 mos.	Iron if anemic; RCT of iron vs. placebo if non-anemic/deficient; placebo if sufficient.	Anemia Hct <33; Iron deficient EP >35.	BSID Stanford-Binet	a) no pretreatment differences any measures b) non-anemic group showed significant improvement in outcome measures at 3 and 6 mos.; both anemic and non-anemic iron deficient groups did not show improvement (regardless of iron therapy), leading to differences at 3 and 6 months c) control group showed greater responsiveness on behavior measures at 3 and 6 months	a) awkward analysis and presentation--insufficient number of controls, inadequate matching and failure to use alternative means of analysis (e.g., regression) b) interpretation of results speculative

Soeoentri	Indonesia; school students in rural province with normal growth and no overt diseases; iron deficient (with anemia) (n=70) vs. replete (n=41)	mean age=10 yrs.	RCT oral iron for 3 mos.	Deficient: Hgb <11, transferrin sat <15%; sufficient: Hgb >12, sat >20%	Achievement Test (standardized for use in Indonesia) Raven Matrix IQ tests (non-verbal, standardized) Concentration test	a) no baseline differences in IQ (not tested post-treatment) or concentration; non-anemic children had higher achievement test results at baseline b) anemic children treated with iron had significant improvement in concentration and in achievement test scores; non-anemic children did not significantly benefit from iron therapy	a)careful analysis--includes pretreatment results. b)SES measures reported as identical between groups; unclear if actually included in analyses. c)generalizability to US situation of milder iron deficiency uncertain, but provides strongest evidence of biologic effect.
Lozoff	Costa Rica, middle class no major medical or developmental problems 191 children total	12-23 months	RCT IM iron vs. po iron vs. placebo	non anemic (Hgb>12-- sufficient depleted deficient intermediate (Hgb10.5-11.9) deficient anemic (Hgb<10.5) deficient	BSID at 1 wk and 3 months	a) at baseline: MDI lower in anemic group only if hgb <10. b) at baseline, psychomotor index lower in anemic vs. non-anemic group c) no effect any therapy at one week d) no overall effect of iron on MDI in anemic group at 3 mos., and severely anemic group (hgb<10) overall remained lower than non-anemic group at end of therapy. Those in severe anemic group who had biochemical evidence of iron repletion were not significantly different from non-anemic group at end of therapy, but this resulted from declining performance of non-anemic group rather than improvement in anemic group.	a) results were not presented for the groups as randomized--authors don't directly answer question of whether providing therapy to anemic, iron deficient children (as would be done in a clinical setting) influences developmental outcome. b) SES differences well controlled--e.g., differences existed between anemic and non-anemic children in "adequacy of home environment" (using HOME inventory), but controlling for these differences did not affect results.

REFERENCES

Alberti PW, Hyde ML, Riko K, Corbin H, Fitzhardinge P, "Issues in Early Identification of Hearing Loss," Laryngoscope 1985;95:373-381.

Alexander, CS, Klassen, AC, "Curtailment of Well Child Services by A Local Health Department: Impact on Rural 2-Year-Olds," Public Health Reports 1986; 101:301-308.

Allen, DB, Bergman, AB, "Social Learning Approaches to Health Education: Utilization of Infant Auto Restraint Devices," Pediatrics 1976; 58:323-328.

Alpert, JJ, Robertson, LS, Kosa, JK, Heagarty, MC, Haggerty, RJ, "Delivery of Health Care for Children: Report of an Experiment," Pediatrics 1976; 57:917-930.

American Academy of Pediatrics, Guidelines for Health Supervision, 1985.

American Academy of Pediatrics, Standards of Child Health Care, Second Edition, 1972.

Anderson, FP, "Evaluation of the Routine Physical Examination of Infants in the First Year of Life," Pediatrics 1970; 45:950-964.

Appelbaum, AS, "Validity of the Revised Denver Developmental Screening Test for Referred and Nonreferred Samples," Psychological Reports 1978; 43:227-233.

Augustin, MS, Stevens, E, Hicks, D, "An Evaluation of the Effectiveness of a Children and Youth Project,", Health Services Report 1973; 88:942-946.

Bass, LW, Wilson, TR, "The Pediatrician's Influence In Private Practice Measured By A Controlled Seat Belt Study," Pediatrics 1964; 700-704.

Bergman, AB, "Use of Education in Preventing Injuries," Pediatric Clinics of North America 1982; 29:331-338.

Berwick, DM, Komaroff, AL, "Cost Effectiveness of Lead Screening," The New England Journal of Medicine 1982; 306:1392-1398.

Bierman, JM, Connor, A, Vaage, M, and Honzik, MP, "Pediatricians' assessment of intelligence of two year olds and their mental test scores," Pediatrics 1964; 34:680.

Black, DAK, "Inequalities in health. Report of a research working group." London: DHSS, 1980.

Black FO, Bergstrom L, Downs M, and Hemenway W., Congenital Deafness Boulder, CO, 1971, cited in Northern J, "Auditory Screening," in Frankenburg WK, Camp BW (eds) Pediatric Screening Tests Springfield, IL, 1975.

Brenner, MH, "Estimating the Effects of Economic Change on National Health and Social Well-Being," Subcommittee on Economic Goals and Intergovernmental Policy of the Joint Economic Committee, Congress of the United States, 1984.

Briscoe, ME, Hochstrasser, DL, Somes, GW, Cowen, DL, Culley, GA, "Follow-up Study of the Impact of a Rural Preventive Care Outreach Program on Children's Health and Use of Medical Services," American Journal of Public Health 1980; 70:151-156.

Brown MJK, Richards SH, and Ambergaokar AG, "Grommets and Ear Glue: A Five Year Followup of a Controlled Trial," J Roy Soc Med 1978;71: 353-6.

Cadman, D, Chambers, LW, Walter, SD, Ferguson, R, Johnston, N, McNamee, J. "Evaluation of Public Health Preschool Child Developmental Screening: The Process and Outcomes of a Community Program," American Journal of Public Health 1987; 77:45-51.

Cadman, D, Chambers, LW, Walter, SD, Feldman, W, Smith, K, and Ferguson, R, "The Usefulness of the Denver Developmental Screening Test to Predict Kindergarten Problems in a General Community Population," American Journal of Public Health 1984; 74:1093-1097.

Camp, BW, van Doorninck, WJ, Frankenburg, WK, Lampe, JM, "Preschool Developmental Testing in Prediction of School Problems," Clinical Pediatrics 1977; 16:257-263.

Canadian Task Force on the Periodic Health Examination. "The periodic health examination," Can Med Assoc J 1979; 121:1193-1254.

Canadian Pediatric Society, "Child Health Care Guidelines," March, 1983.

Cantekin EI, Mandel EM, Bluestone CD, et al., "Lack of Efficacy of a Decongestant-Antihistamine Combination for Otitis Media in Children," NEJM 1983; 308:297-301.

Casey, P, Sharp, M, Loda, F, "Child-Health Supervision for children under 2 years of age: A review of its content and effectiveness." The Journal of Pediatrics 1979; 95:1-9.

Casey, PH, Whitt, JK, "Effect of the Pediatrician on the Mother-Infant Relationship," Pediatrics 1980; 65:815-820.

Casey, PH, Bradley, RH, Caldwell, BM, Edwards, DR, "Developmental Intervention: A Pediatric Clinical Review," Pediatric Clinics of North America 1986; 33:899-923.

Chamberlin, RW, Szumowski, BA, "A Follow-Up Study of Parent Education in Pediatric Office Practices: Impact at Age Two and a Half," AJPH 1980; 70: 1180-1188.

Chamberlin, RW, "Developmental Assessment and Early Intervention Programs for Young Children: Lessons Learned From Longitudinal Research," Pediatrics in Reveiw 1978; 8

Charney, E, "Well-Child Care as Axiom," in Charney, E, ed., Well Child Care, Report of the Seventeenth Ross Roundtable on Critical Approaches to Common Pediatric problems. Columbus, Ohio: Ross Laboratories, 1986, p. 1-8.

Christophersen, ER, "Accident Prevention in Primary Care," Pediatric Clinics of North America 1986; 33: 925-933.

Committee on Standards of Child Health Care (American Academy of Pediatrics), Standards of Child Health Care (3rd edition) Evanston, Illinois, 1977, pp. 138-48.

Coplan J, "Deafness: Ever Hear of It? Delayed Recognition of Permanent Hearing Loss," Pediatrics 1987; 79:206-213

Cullen, KJ, "A six-year controlled trial of prevention of children's behavior disorders," The Journal of Pediatrics 1976; 88:662-666.

Dawson, P, Cohrs, M, Eversole, C, Frankenburg, WK, Roth, ML, "Cost-Effectiveness of Screening Children in Housing Projects," AJPH 1976; 66:1192-1194.

DeAngelis, C, Berman, B, Oda, D, Meeker, R, "Comparative values of school physical examinations and mass screening tests," The Journal of Pediatrics 1983; 102:477-481.

Deinard, A, Gilbert, A, Dodds, M, and Egeland, B, "Iron Deficiency and Behavioral Deficits," Pediatrics 1981; 68:828-833.

Deinard, AS, List, A, Lindgren, B, Hunt, JV, Chang, P, "Cognitive deficits in iron-deficient and iron-deficient anemic children, The Journal of Pediatrics 1986; 108:681-689.

Driggers, DA, Reeves, JD, Lo, EYL, and Dallman, P, "Iron Deficiency in one year old infants: Comparison of results of a therapeutic trial in infants with anemia or low-normal hemoglobin values," Journal of Pediatrics 1981; 98:753-758.

Dutton, DB, Silber, RS, "Children's Health Outcomes in Six Different Ambulatory Care Delivery Systems", Medical Care 1980; 18:693-714.

Feldman W, Milner RA, Backett B, and Gilbert S, "Effects of Preschool Screening for Vision and Hearing on Prevalence of Vision and Hearing Problems 6-12 Months Later," Lancet 1980; 2:1014

Fisch L, "Development of School Screening Audiometry," British Journal of Audiology 1981; 15: 87-95.

FitzZaland RE and Zink GD, "A Comparative Study of Hearing Screening Procedures," Ear and Hearing 1984; 5:205-210.

Foxman, B., Lohr, KN, Brook, RH, Measurement of Physiologic Health for Children Vol 5: Anemia Santa Monica, CA, 1983.

Frankenburg, WK, "Developmental Assessment," in Levine, MD, Carey, WB, Crocker, AC, and Gross, RT, eds., Developmental and Behavioral Pediatrics,Philadelphia, 1983, pp. 927-937.

Frankenburg, WK, Goldstein, AD, Camp, BW, "The revised Denver Developmental Screening Test: Its accuracy as a screening instrument," The Journal of Pediatrics 1971; 79:988-995.

Frankenburg, WK, Dodds, JB, "The Denver Developmental Screening Test," The Journal of Pediatrics 1967; 71:181-191.

Gilbert, JR, Feldman, W, Seigal, Linda, Mills, DA, Dunnett, C, and Stoddart, G, "How many well-baby visits are necessary in the first 2 years of life?" Can Med Assoc J 1984; 130:857-861.

Gordis, L, "Effectiveness of Comprehensive-Care Programs in Preventing Rheumatic Fever," The New England Journal of Medicine 1973: 289:331- 335.

Gordis, L, Markowictz, M, "Evaluation of the Effectiveness of Comprehensive and Continuous Pediatric Care," Pediatrics 1971; 48:766.

Grant, WW, Fearnow, RG, Hebertson, LM, Henderson, AL, "Health Screening in School-Age Children," Am J Dis Child 1973, 125:520-522.

Great Britain, Committee on Child Health Services (Court, SDM, chair), Fit for the Future HMSO, London, 1976.

Gutelius, MF, Kirsch, AD, MacDonald, S, Brooks, MR, McErlean, T. "Controlled Study of Child Health Supervision: Behavioral Results," Pediatrics 1977; 60:294-304.

Haggerty, R, "The Rand health insurance experiment for children, commentary," Pediatrics 1985; 76:969-971.

Hall, Dr. David, Personal Communication, February 9, 1987.

Harper, DC, Wacker, DP, "The Efficiency of the Denver Developmental Screening Test with Rural Disadvantaged Preschool Children," Journal of Pediatric Psychology 1983; 8:273-283.

Hoekelman, RA, "What Constitutes Adequate Well-Baby Care," Pediatrics 1975; 55:313-325.

Hoekelman, RA, "Well Child Visits Revisited," American Journal of Diseases of Children 1983; 137:17-20.

Hutton, Nancy, and Starfield, B, "Iron Deficiency Anemia," in Starfield, ed., The Effectiveness of Medical Care: Validating Clinical Wisdom Baltimore, 1985.

Irwin, PH, Conroy-Hughes, R, "EPSDT Impact on Health Status: Estimates Based on Secondary Analysis of Administratively Generated Data," Medical Care 1982; 20: 216-234.

Joint Committee on Infant Hearing," Position Statement," Pediatrics 1982; 70:496-7.

Kanthor, HA, "Car Safety for Infants: Effectiveness of Prenatal Counseling", Pediatrics 1976; 58:320-322.

Kaplan, RS, Lave, LB, Leinhardt, S, "The Efficacy of a Comprehensive Health Care Project: An Empirical Analysis," American Journal of Public Health 1972; 62:924-930.

Keller, W, "Study of Selected Outcomes of the Early and Periodic Screening, Diagnosis, and Treatment Program in Michigan," Public Health Reports 1983; 98: 110-119.

Kelly B, Sein C, McCarthy P, "Safety Education in a Pediatrics Primary Care Setting," Pediatrics 1987; 79: 818-24.

Kessner, DM, Snow, CK, Singer, J, Assessment of Medical Care for Children National Academy of Sciences, Washington, DC, 1974.

Kessner, J, and Kalk, L, A Strategy for evaluating health services Washington, D.C., 1973.

Kim, I, Pollitt, E, Leibel, RL, Viteri, FE, Alvarez, E, "Application of Receiver-Operator Analysis to Diagnostic Tests of Iron Deficiency in Man," Pediatric Research 1984; pp. 916-920.

Klein, M, Roghmann, K, Woodward, K, and Charney, E, "The Impact of the Rochester Neighborhood Health Center on Hospitalization of Children, 1968 to 1970," Pediatrics 1973; 51:833-839.

Kohler, L. "Health Control of Four-year Old Children." Acta Pediatr Scand 1973; suppl 235, 1-28.

Kohler, L, Holst, H-E, "Auditory Screening of Four-Year Old Children," Acta Paediatr Scand 1972; 61:555-60.

Kohler, L, "Physical Mass Examinations in the School Health Service," Acta Paediatr Scand 1977; 66:307-310.

Korsch, B, Cobb, K, Asche, B, "Pediatricians' appraisals of patients' intelligence," Pediatrics 1961; 27:990.

Korsch, BM, "Issues in evaluating child health supervision," The Journal of Pediatrics 1985; 75:942-951.

Lanson, P, "Efficacy of Prenatal and Postpartum Home Visits on Child Health and Development," Pediatrics 1980; 66:191-197.

Leibowitz, A, Manning WG, Keeler, EB, et al, "Effect of cost sharing on the use of medical services by children: Interim results from a randomized controlled trial," Pediatrics 1985; 75: 942-951.

Lewis, C, "What Is the Evidence?" Amer J Dis Child 1971; 469-474.

Lobovits A. "Auditory Screening" Advances in Pediatrics 1984; 420.

Lozoff B, Birttenham GM, Wolf AW, et al., "Iron Deficiency Anemia and Iron Therapy Effects on Infant Developmental Test Performance," Pediatrics 1987; 79:981-995.

Lozoff, B, Brittenha, GM, Viteri, FE, Wolf, AW, Urrutia, JJ, "Developmental deficits in iron-deficient infants: Effects of age and severity of iron lack," The Journal of Pediatrics 1982;101:948-951.

Lozoff, B, Brittenham, GM, Viteri, FE, Wolf, AW, Urrutia, JJ, "The effects of short-term oral iron therapy on developmental deficits in iron-deficient anemic infants," The Journal of Pediatrics 1982; 100:351-357.

Macfarlane, Dr. Aidan, Personal Communication, February 18, 1987.

Mandel EM, Rockette HE, Bluestone CD, Paradise JL, and Nozza R,"Efficacy of Amoxicillin with and without Decongestant-Antihistamine for Otitis Media with Effusion in Children," NEJM 1987;316:432-7.

Marchessault, JHV M.D., FRCPC, "Personal Communication," February 18, 1987.

Meisels, SJ, Margolis, LH, Is the Early and Periodic Screening, Diagnosis, and Treatment Program Effective With Developmentally Disabled Children? Pediatrics. 1988;81: 262-271.

Miller, JR, Pless, IB, "Child Automobile Restraints: Evaluation of Health Education," Pediatrics 1977; 59:907-911.

Moore, G, Frank, K, "Comprehensive Health Services for Children: An Exploratory Study of Benefit," Pediatrics 1973; 51:17-21.

Nader, P, Roy, L, and Briak, S, "The New Morbidity: Use of School and Community Health Care Resources for Behavioral, Educational and Social Family Problems," Pediatrics 1981; 67:53-59.

Nader, P, "Frequency and nature of school problems," in Haggerty, R, Roughman, and Pless, IB, eds., Child Health and The Community New York, 1975.

Northern J, "Auditory Screening," in Frankenburg, Camp ed. Pediatric Screening Tests, Springfield, IL, 1975.

Northern J, "Impedence Audiometry," in Northern, J, ed., Hearing Disorders (2nd Edition) Boston, 1984, pp. 41-56.

O'Connell, EJ, Friesen, CD, "The Preschool Physical Examination," Clinical Pediatrics 1976; 15:930-931.

Olds, DL, Henderson, CR, Tatelbaum, R, Chamberlin, R, "Improving the Delivery of Prenatal Care and Outcomes of Pregnancy: A Randomized Trial of Nurse Home Visitation," Pediatrics 1986; 77:16-28.

Osborne L, Wooley FR, "Use of Groups in Well Child Care," Pediatrics 1981; 67:701.

Oski, FA, Honig, AS, Helu, B, Howanitz, P, "Effect of Iron Therapy on Behavior Performance in Nonanemic, Iron-Deficient Infants," Pediatrics 1983; 71:877-880.

Oski, FA, Honig, AS, "The effects of therapy on the developmental scores of iron-deficient infants," The Journal of Pediatrics 1978; 92:21-25.

Paradise JL, "Otitis Media During Early Life: How Hazardous to Development? A Critical Review of the Evidence," Pediatrics 1981;68:869-73.

Pettigrew AG, Edwards DA, and Henderson-Smart DJ, "Screening for Auditory Dysfunction in High Risk Neonates," Early Human Development 1986; 14: 109-20.

Pless, IB, "Accident Prevention and Health Education: Back to the Drawing Board?," Pediatrics 1978; 62:431-435.

Pollitt, E, Leibel, RL, "Iron deficiency and behavior," The Journal of Pediatrics 1976; 88:372-381

Ramey, CT, Sparling, JJ, Bryant, DM, Wasik, BH, "Primary Prevention of Developmental Retardation During Infancy," Prev. Human Services 1982; 1:61-83.

Reeves, JD, Bicninsky, E, Addiego, J, Lubin, BH, "Iron Deficiency in Health and Disease," Advances in Pediatrics 1984; 281-319.

Reis, JS, Pliska, SR, Hughes, EFX, "A Synopsis of Federal-State Sponsored Preventive Child Health," Journal of Community Health 1984; 9: 222-239.

Reisinger, KS, Williams, AF, Wells, JK, John, CE, Roberts, TR, Podgainy, HJ, "Effects of Pediatricians' Counseling on Infant Restraint Use," Pediatrics 1981; 67:201-206.

Reisinger, KS, Williams, AF, "Evaluation of Programs Designed To Increase the Protection of Infants in Cars," Pediatrics 1978; 62: 280-287.

Reisinger, KS, Bires, JA, "Anticipatory Guidance in Pediatric Practice," Pediatrics 1980; 66:889-892.

Rogers, KD, Ernst, R, Shulman, I, Reisinger, KS, "Effectiveness of Aggressive Follow-up on Navajo Infant Health and Medical Care Use," Pediatrics 1974; 53:721-725.

Rosenbaum, S, & Johnson, K, "Providing Health Care for Low-income Children: Reconciling Child Health Goals with Child Health Financing Realities," Milbank Quarterly 1986; 64:442-478.

Roughmann, KJ, "Intervention Strategies for Children: A Research Agenda," Health Services Research 1985; 19:6 887-943

Sackett, DL, Haynes, RB, Tugwell, P, Clinical Epidemiology--A Basic Science for Clinical Medicine Boston, 1985.

Scherz, RG, "Restraint Systems for the Prevention Of Injury to Children in Automobile Accidents," AJPH 1976; 66:451-456.

Shadish, WR, "A Review and Critique of Controlled Studies of the Effectiveness of Preventive Child Health Care," Health Policy Quarterly 1982; 2:24-52.

Soemantri, AG, Pollitt, E, Kim, I, "Iron deficiency anemia and educational achievement," The American Journal of Clinical Nutrition 1985; 42:1221-1228.

Starfield, B, Dutton, D, "Care, Costs, and Health: Reactions to and Reinterpretation of the Rand Findings," Pediatrics 1985;76:614-621.

Strain, JE, "AAP Periodicity Guidelines: A Framework for Educating Patients," Pediatrics 1983; 74 (Supplement): 924-927.

Sturner, RA, Green, JA, Funk, SG, "Preschool Denver Developmental Screening Test as a predictor of later school problems," The Journal of Pediatrics 1985; 107:615-621.

The General Medical Services Committee of the British Medical Association and the Royal College of General Practitioners, Handbook of Preventive Care for Pre-school Children London, 1984.

Thomas KA, Hassanein RS, and Christopherson ER, "Evaluation of Group Well Child Care for Improving Burn Prevention in the Home," Pediatrics 1984; 74: 879-82.

Valdez, ROB, The Effects of Cost Sharing on the Health of Children Rand. Santa Monica, CA, R-3270-HHS, 1986.

Vazquez-Sloane P, Windom R, and Pearson HA, "Disappearance of Iron-Deficiency Anemia in High-Risk Population Given Supplemental Iron, " N Engl Jour Med 1987; 313:1239-40.

Ventry IM, "Effects of Conductive Hearing Loss: Fact or Fiction?," J Speech Hear Disorder 1980;45:143-56.

Walter, T, Kovalskys, J, Stekel, A, "Effect of mild iron deficiency on infant mental development scores," The Journal of Pediatrics 1983; 102:519-522.

Webb, TE, Oski, FA, "Iron deficiency anemia and scholastic achievement in young adolescents," The Journal of Pediatrics 1983; 82:827- 830.

Welch, NM, Kesler, RW, "The value of the preschool examination in screening for health problems," The Journal of Pediatrics 1982; 100:232-234.

Yankauer, A, Lawrence, RA, "A Study of Periodic School Medical Examinations--I. Methodology and Initial Findings," AJPH 1955; 45:71-78.

Yankauer, A, Lawrence, R, Ballou, L, "A Study of Periodic School Medical Examinations--III. The Remediability of Certain Categories of "Defects", AJPH 1957; 47:1421-1429.

Yankauer, A, Lawrence, R, "A Study of Periodic School Medical Examinations--II. The Annual Increment of New "Defects"," AJPH 1956; 55:1553-1562.

Yankauer, A, "Child Health Supervision--Is It Worth It?" Pediatrics 1973; 52:272-279.